DOCTEUR EN SCIENCES ÉCONOMIQUES ET DE GESTION

Essays on Socio-Economic Determinants of Risky Sexual Behavior and HIV/AIDS: Insights from Sub-Sharan Africa

Tequame, Miron

Award date:
2013

Awarding institution:
Universite de Namur

University of Namur

Membre de l'Académie universitaire « Louvain »

Faculté des Sciences Economiques, Sociales et de Gestion

Département des Sciences Economiques

Dissertation pour l'obtention du titre de

DOCTEUR EN SCIENCES ECONOMIQUES ET DE GESTION

Essays on Socio-Economic Determinants of Risky Sexual Behavior and HIV: Insights from Sub-Saharan Africa

Miron Tequame

Octobre 2013

Composition du Jury

Promoteur : Prof. Jean-Marie Baland (secrétaire)

Autres membres du Jury : Damien de Walque (World Bank)
Menno Pradhan (Free University Amsterdam)
Prof. Jean-Philippe Platteau (UNamur)

Président du Jury : Prof. Gani Aldashev

Acknowledgements:

The PhD in Namur is among the most difficult challenges I took in my life. It is with great relief that I reached to this end. Of course nothing would have been possible if not for many nice people I encountered during my trip. I am grateful to Jean Marie Baland who has always been very nice and supportive in very difficult moments. It is thanks to his patience that I am able to reach where I am. I am also grateful to my co-Supervisor Gani Aldashev and Jean Philippe Platteau for their indirect support and guidance during my stay at CRED. An immense thankfulness goes to the other members of the jury, Damien de Walque and Menno Pradhan who transformed my dissertation with their great comments. I am also particularly grateful to Cathrine Guirkinger for the time she took to provide comments and her guidance while writing my dissertation.

I would also like to take the time and thank all the other members of CRED. I have encountered many wonderful people at CRED and it would be difficult to list all of them. Some of them have been really incredible in dragging me back to track during difficult moments. I deeply thank Petros and Michel who always encouraged me to see the light when all seemed dark.

This PhD belongs to my father to whom I am deeply indebted for his persistence in encouraging me all the time and whose expectations were always high to keep me focused on my objectives. Of course nothing would have been possible without the support of Marco who has been a family during all these years. My mothers and my sisters are other wonder wonderful people to whom I will always be grateful.

Contents

1 Introduction . . . 6
1.1 HIV, Risky Sexual Behavior and Ethno-Linguistic Heterogeneity . . . 7
1.2 Pregnancy Response when Scaling-Up Anti-Retroviral Therapy . . . 8
1.3 Siblings Sex Composition and Risky Sexual Behavior . . . 8
1.4 Thesis Organization . . . 8
2 HIV, Risky Sexual Behavior and Ethno-Linguistic Heterogeneity . . . 10
2.1 Conceptual Framework . . . 12
2.2 Empirical Strategy and Results . . . 15
2.3 Discordant Couples and Ethno-Linguistic Fractionalization . . . 24
2.4 Sensitivity Analysis . . . 26
2.5 Conclusions . . . 33
3 Free Access to HAART and Pregnancy Response among HIV Patients. A Case Study from Cameroon . . . 35
3.1 Data . . . 38
3.2 Empirical Strategy . . . 43
3.3 Impact of Treatment on Pregnancy . . . 45
3.4 Heterogeneous Treatment Effects . . . 50
3.5 Threats to Internal Validity . . . 54
3.6 Discussion and Concluding Remarks . . . 57
4 Sex Composition of Older Siblings and Teen-Pregnancy in Patriarchal Communities. Evidence from Central Africa . . . 60
4.1 Age and Gender Based Roles among Teen Siblings in SSA Context . . . 63
4.2 Data and Empirical Strategy . . . 66
4.3 Results . . . 75
4.4 Mechanisms and Alternative Explanations . . . 78
4.5 Welfare Consequences of Premarital Pregnancy while Teen . . . 86
4.6 Conclusion . . . 90

List of Figures

1 HIV Prevalence by Compliance Quintiles. 29
2 HIV Prevalence by ELF Quantiles 29
3 HIV Compliance by ELF Quantiles 29
4 Birth Trends: Total, Intended and Unintended 42
5 Knowledge of Women on their Partner's Status 52
6 Lowess Estimate of CD4 at Treatment Initiation 52
7 Individuals on Treatment Before 2006 56
8 Age at First Birth: Within/Out-of-Wedlock 66
9 Kernel Distribution of Age at First Premarital Birth 67
10 Distribution of Male Firstborns 71
11 Cox Hazard Estimates on Secondborn Women aged between 15-24 78

List of Tables

1	Descriptive Statistics	16
2	Ethno-linguistic Heterogeneity on HIV status and HIV prevalence	20
3	Belonging to a Minority/Majority group on HIV status	20
4	The Effect of Ethno-linguistic Heterogeneity on HIV Status by Gender	22
5	HIV Status and Different indicators of Ethno-linguistic Heterogeneity	23
6	HIV Status and Discordance Among Couples	25
7	Extramarital Sex and Ethno-linguistic Heterogeneity	28
8	Socio-Economic Difference between Compliers and Non Compliers	30
9	The Effect of ELF on HIV status with different levels of compliance	32
10	The Effect of ELF on HIV status with different levels of Migration	34
11	Medical Record on date of Treatment Initiation	39
12	Descriptive Statistics by Treatment Year	41
13	Mean Comparison between Treated and Non-Treated on Perception of HIV/AIDS	43
14	Coefficients from DIFFinDIFF based on the Moving Window: Birth in t+1 compared to t	46
15	Dynamic of the Coefficients by Year of Treatment Initiation	47
16	Coefficients Estimates of Treatment on Pregnancy	49
17	Coefficient Estimates by Number of Pre-Treatment Children	50
18	Random Individual Specific Trend dependent on Initial CD-4	53
19	Sensitivity analysis: Attrition, Selection and Omitted Variables	55
20	Descriptive Statistics by Country and Cohorts	68
21	Mean Difference by Gender of Firstborn and Survival Status	73
22	Male Firstborn Effect on Teen Pregnancy of Never Married Teen Secondborns	77
23	Cox Hazard Estimates on Age at First Birth	78
24	Preceding and Succeeding Males on Premarital Pregnancy	80
25	Mechanisms and Siblings Interaction on Premarital Pregnancy	83
26	The Role of Preceding Males and Age Gap with Partner	84

27 Premarital Pregnancy and Marital Status across Cohorts 88
28 Premarital Pregnancy and Marriage Quality across Cohorts 89
29 Premarital Pregnancy and Survival Status of First Child across Cohorts 91

1 Introduction

THIS thesis is composed of three essays on socio-economic determinants of risky sexual behavior and HIV/AIDS in Sub-Saharan Africa. Sub-Saharan Africa accounts for 70% of new HIV infections, being the most severely devastated continent by the epidemics by hosting 68% of approximately 34 million people living with HIV/AIDS worldwide.[1]

Many studies have documented the long-term consequences of HIV/AIDS: a big drop in life expectancy;[2] death of adults during reproductive age and vulnerable orphans (Case *et al.,* 2004); displacement of firms due to morbidity and absenteeism of workers; death of adults in their productive age who leave members of their family in economic crisis (Whiteside 2002) and many more implications. Fortunately, AIDS treatment is increasingly accessible for a larger proportion of the population. However, on average, only one third of HIV infected individuals who are eligible for treatment have access to anti-retroviral therapy (ART).

There is a general consensus that the primary channel of HIV/AIDS transmission in Africa is heterosexual sex (Schmid *et al.*, 2004). Indeed, the DHS Comparative Report (February 2009) shows that HIV infection is associated with an increasing number of lifetime sex-partners, earlier age of sexual debut among women, sex with non-marital or non-cohabiting partners and alcohol use during last sex. By now, there is also a growing body of research focusing on socio-economic determinants of risky sexual behavior and hence HIV/AIDS in Africa. Among the many characteristics outlined are: poverty, migration, cost of HIV testing, circumcision, sugar daddies and political instability. (Decosas *et al.,* 1995, Dupas 2011, Dupas *et al.,* 2011, Dupas *et al.,* 2010 , Thornton 2008 , Gray *et al.*, 2007, Whitside 2002[120]).

This research contributes to the framework of the above literature by highlighting community or individual level characteristics that shape risky sexual behavior. The approach is based on the discussion and provision of literature reviews on mechanisms that drive changes in risky behavior. These diverse mechanisms are supported with empirical analysis from two sources of data: the Demographic and Health Survey and firsthand data collected in Cameroon. The thesis starts by looking at how community level characteristics like ethno-linguistic heterogeneity shape individuals' risky sexual behavior. It puts in evidence a peculiar aspect of African societies charac-

[1] UNAIDS World AIDS day report, 2011

[2] 2009 AIDS Epidemic Update stated that life expectancy has dropped on average to 45 in countries most affected by HIV.

terized by ethno-linguistic diversity. I argue, in line with the literature, different forms of heterogeneity reduce social networks and information on deviating behaviors like extramarital sex; such environments are favorable specifically for women who want to engage in risky sexual behavior. The second essay looks at pregnancy response of women when treatment initiating by using firsthand data from Cameroon. Certain societies value high number of children as they represent richness. This might induce individuals to engage in risky sexual behavior intentionally for reproductive reasons and avoid different forms of stigma. Finally, I look at family structure and sex composition of siblings. In African contest, where sex and age-based hierarchy among siblings is widespread, the study investigates how preceding sex composition of siblings affects teen pregnancy. The idea behind is gender and age create different forms of authority among members of the family which result in monitoring behaviors among younger teen girls. In the following paragraphs, an abstract of each essay, single authored, is outlined for illustrative and briefing purpose.

1.1 HIV, Risky Sexual Behavior and Ethno-Linguistic Heterogeneity

Ethno-linguistic heterogeneity is associated with indicators of development like civil society, trust, quality of institutions, economic performance and participation. Recently, it has been found to be favorable for optimizing agents who want to engage in risky sexual behavior as they can select partners outside their own network and that of their spouse (Pongou 2011). This paper augments the literature by arguing that the effect of ethno-linguistic heterogeneity on risky sexual behavior is gender related. In contexts where women have more economic benefits (or losses) from sexual partners and their extramarital sexual relations are highly stigmatized, they better take advantage from the reduced likelihood of detection. With respect to men, ethno-linguistic heterogeneity significantly affects women's HIV status and their risky sexual behavior. Accordingly, it affects positively the probability of being in a discordant couple where the wife is HIV positive. The analysis is based on empirical evidence from the DHS surveys in Sub-Saharan Countries.

1.2 Pregnancy Response when Scaling-Up Anti-Retroviral Therapy

The HIV/AIDS epidemic has dramatically altered patterns of morbidity and mortality with different consequences on fertility behavior. The study takes advantage of a unique data-set collected in Cameroon among HIV positive patients and estimates the relationship between HAART treatment and (intended) pregnancy. The direct health benefits of treatment imply rational and behavioral response in pregnancy as it allows individuals to accomplish their desired number of children. A before-after analysis is conducted to evaluate the effect of the 2007 policy based on scaling-up HAART treatment at national level. With respect to women not yet on treatment, HAART increased the propensity to pregnancy after one year with the coefficient increasing over time after 2007. Pregnancy response was highest among people who have lower number of children pre-treatment and with CD4 counts above the average at treatment initiation. The study discusses and tests different mechanisms that drive the behavioral response in Yaounde-Cameroon.

1.3 Siblings Sex Composition and Risky Sexual Behavior

Sex and birth order of siblings are found to be important determinants of economic outcomes and future well-being. This study augments the literature by examining how gender of the firstborn matters in shaping risky sexual behavior of the younger ones in Central Africa. Based on the DHS survey, it shows that male firstborns, who enjoy authority due to pro-male biased society, better shape risky sexual behavior of younger teen sisters as compared to female firstborns. This mechanism holds for any preceding male where age-based hierarchy interacted with gender creates positive externality for the household by reducing the associated cost of premarital teen pregnancy. The study puts in evidence the mechanisms and illustrates the welfare consequences of premarital teen pregnancy.

1.4 Thesis Organization

The structure of the thesis is as follows:

Chapter 2 presents the research on HIV, Ethno-linguistic Heterogeneity and Risky Sexual Behavior. The section is composed of an overview, theoretical motivation, data and methodology. Finally, main conclusions are discussed in the last section.

Chapter 3 outlines the paper on Pregnancy Response of Women Living with HIV/AIDS with Scaling-up of HAART Treatment based on data collected in Cameroon. The section gives an overview of the existing literature, the empirical strategy used with description of the data, treats to internal validity and finally conclusions that can be drown from data analysis.

Chapter 4 , on Gender of Older Siblings and Teen-Pregnancy, presents some insights on the role of hierarchy and gender among siblings in Central Africa. It starts by illustrating the role of teen pregnancy, the associated costs and how parents coop with it. The section describes the data and identification strategy and conclusive remarks. Welfare consequences of teen pregnancy are also discussed.

2 HIV, Risky Sexual Behavior and Ethno-Linguistic Heterogeneity

Nowadays, risky sexual behavior is a major focus of HIV prevention policy in Africa. Understanding the behavioral response in the region is an important base for predicting the future path of the epidemic. Studies in Africa have shown that general response in risky sexual behavior due to the epidemics is ambiguous (Stoneburner and Low-Beer, 2004; Bloom *et al.*, 2000; Williams *et al.*, 2003). It is found to be heterogeneous based on socio-economic characteristics like education of women (De Walque, 2007; Dupas, 2009), future life expectancies and expected revenues (Oster, 2012). The anti-HIV education labeled ABC-Abstain, Be faithful and use Condom- has proven to be effective in Uganda (Green et al, 2006) and it has been extended to other sub-Saharan countries without, up to now, any clear evidence of success.[3]

Other than behavioral factors at the individual level, treatment of non-HIV sexually transmitted diseases, including circumcision of men, has proven to be effective in preventing HIV/AIDS (Weiss *et al.*, 2000) and shaping its regional pattern.[4] Also, local economic development plays a role on HIV infection through transactional sex, as emphasized in Oster (2007). Under the assumption that risky sexual behavior is a normal good, economic development, induced by exports, affects individuals' HIV status.

In contrast to large number of studies examining individual-level features and determinants of infection, few studies have focused on community level characteristics and networks that affect individual's sexual behavior. This research contributes in analyzing the diverse rates of HIV prevalence across regions and gives insights in understanding the role ethno-linguistic heterogeneity has on risky sexual behavior. Heterogeneity in terms of ethnicity, race or religion has important implications on social capital and, consequently, on economic variables.[5] Such communities give little space for informal social networks, making the flow of information on individual's deviating behaviors less likely to be detected. It is also associated with lower trust

[3] However, Oster (2007) gives alternative possible reasons why ABC was claimed to be effective in Uganda. She explains that routes of exports have an important role rather than population based educational campaign.

[4] The high rate of HIV transmission in Africa is due to other untreated sexually transmitted diseases (STD). The difference in transmission rates is large enough to explain the observed difference in prevalence between the United States and Sub-Saharan Africa (Oster 2005).

[5] For example, Alesina and La Ferrara (2000) found that ethnic, racial and economic heterogeneity have a negative impact on social capital and consequently on economic development in the USA.

among communities because members of the community have different tastes making it difficult to enforce a system of social sanctions and cohesion. Such types of communities leave less-or-unsanctioned socially deviating behaviors.

Pongou (2011) is the first paper to show that individuals living in heterogeneous societies are more likely to engage in extramarital sex because they can easily avoid being detected by their partner. The paper developed a theory to explain how community-level ethnic heterogeneity determines the formation of sexual networks among couples, and how this, in turn, affects the spread of HIV/AIDS. In the model, agents derive utility from sexual relationships and infidelity is punished if detected by own partner. When information circulates more easily within ethnic groups than across, agents tend to choose their extramarital partners from different groups to hide their infidelity, thanks to cross-group anonymity.

This chapter extends the findings of Pongou (2011) by underlining the role of gender-gap in extramarital sex. In many communities cost of detection from extramarital sex is higher for women for two reasons. Firstly, gender specific roles in sexuality require only men to be risk takers with multiple partners, defined as masculinity.[6] Furthermore, in a gender biased economy where individuals gain non-material and material benefits from engaging in extramarital sex, the cost of being sanctioned from detection is augmented for women by the material loss incurred from their partner. Thus women internalize better community characteristics and ethno-linguistic heterogeneity when optimizing the number of extramarital sex. With respect to Pongou (2011), this study highlights a particular aspect of several African societies: pro-male economy and patriarchal tradition. Both characteristics make women's adultery highly stigmatized and costly. I argue that ethno-linguistic heterogeneity shapes more women's behavior rather than men's by giving evidence of higher elasticity for women. The extension is further supported by the analysis conducted on couples' data-set and concordance of their HIV status.

I use the Demographic and Health Survey data-sets (DHS) to link HIV status of individuals with their ethnic characteristic. The study is focused on four sub-Saharan countries: Malawi, Cameroon, Kenya and Ethiopia in order of HIV prevalence. The choice of these countries is driven by data availability on ethnicity, territorial comparability and relevant HIV prevalence.[7] The period of survey ranges between 2003 and

[6] Men tend to have/declare more extramarital sex then women. For a literature review on gender differences in sexuality refer to Oliver *et al.* (1993).

[7] With respect to Pongou (2011) I exclude Burkina Faso and Ghana for territorial comparability. HIV

2005.

The variable of interest is computed with three different measures of heterogeneity widely used in the literature: the Ethno-Linguistic Fractionalization (ELF) index, based on the Herfindahl index; Generalized Ethno-Linguistic Fractionalization (GELF), an extension of the ELF which takes into account similarity between ethnic groups based on cross-ethnic marriages; the Entropy index (EI) which is an indicator of diversity. The three measures differ for the degree of interaction among ethnic groups as well as the number of ethnicities in the community. The ELF gives more weight to individuals with higher proportion in the population. The GELF takes into account some aspects of similarity between the groups. The Entropy index, on the contrary, gives more weight to the number of ethnicities in the community.

Results suggest ethno-linguistic heterogeneity is more elastic to HIV status of women rather than men. This finding is not driven by the biological exposure of women to HIV infection. Notwithstanding, I further analyze the data-set of couples and their HIV status by focusing on discordant couples who are married for more than 5 years. I find a positive relationship between the indicators and the probability of being in a discordant couple where she is positive and the husband is negative. On the reverse, I do not find the same relationship on discordant couples where he is positive. This finding suggests women are more sensitive to social environment when engaging in risky sexual behavior.

The chapter is organized as follows: Section 2.1 gives an outline of the mechanism that drive risky behavior; 2.2 illustrates the identification strategy and results. Sensitivity analysis is conducted in Section 2.4 and finally, some conclusions are laid off in Section 2.5.

2.1 Conceptual Framework

Ethnic fractionalization has an important role in the political economy of many countries as it leads to political instability, poor quality of institutions, badly designed economic policy and poor economic performance. Several studies have shown the negative relationship between racial or ethnic heterogeneity and the provision of public goods (Alesina *et al.*, 1999; Miguel and Gugerty, 2005a) while others focused on

prevalence is very low in these two countries.

its impact on productivity (Bandiera *et al.*, 2005; Alesina and La Ferrara, 2005). For example, a comparison across US counties showed that higher ethnic fractionalization is associated with a lower rate of collective action in the community (Vigdor, 2004). There are also studies that have linked ethnic fractionalization with lower trust and participation in the communities (Alesina and La Ferrara, 2000).

The role of ethno-linguistic heterogeneity on low public expenditures and provision of public goods has an implication on HIV/AIDS. Indeed, in fragmented societies, governments are less likely to provide anti-HIV policies because of lower civil society and collective action (Lieberman, 2007).

Measuring, and thus defining, ethno-linguistic heterogeneity in sub-Saharan countries has been challenging. The traditional measure based on the Herfindahl index, conventionally called *Ethno-Linguistic Fractionalization* (ELF), is given by the probability of two randomly drawn individuals from a given population belong to two different groups. If I consider a community composed of more than two different ethnic groups, with s_i being the share of group i over the total at community level and n being the total number of ethnicity, then the ELF index would be:

$$ELF = 1 - \sum_{i=1}^{n} s_i^2 \quad (1)$$

It varies between 0-1 and an increase in the indicator means more heterogeneous society. The ELF assumes that language and ethnicity map one-to-one and both have similar consequences on individuals' networking and social sanctions.

Social heterogeneity is endogenous to many unobservables at community level. It is a continuously evolving indicator and difficult to define across different cultural, time and space boundaries. For example, continuous regional migration and inter-mixing with other groups suggests it is a function of local economic development, colonial policy and inter-group mixing (Alesina *et al.*, 2003). More detailed critics to indicators of ethno-linguistic heterogeneity in Africa have been argued in Posner (2004).

In some societies there are different similarities among different ethnic groups which compose the community. The ELF index gives the same weight to all ethnicity, while in reality some might have similar cultural values which should be taken into account. The ELF index attributes a 0 or 1 value to individuals in the community depending if they belong or not to an ethnicity or race. A generalized version of ELF (Generalized Index of Ethno-linguistic Hetergeneity-GELF) has been proposed in Bossert *et al.* (2008) where each individual is weighted by similarity values. Similarity measure-

ment can be based on income, education, language and etc. To compute the indicator of similarity, I use the matching of ethnicities from the data-set of couples. The reasoning behind is if a high proportion of people from ethnicity «x» get married to another ethnicity «y» then similarity between these two ethnicities is non zero in terms of cultural values and networks. To compute similarities between two ethnicities, I take the average of four frequencies: frequency of wife from ethnicity «x» getting married to husband from ethnicity «y»; frequency of wife from ethnicity «y» getting married to husband from ethnicity «x»; and similarly for the husband. To give you some illustration on the indicator I construct, let us consider a community with three ethnicities. The similarity matrix, S, of couples' matching based on the ethnicity of the husband and wife is as follows:

$$S = \begin{bmatrix} 1 & s_{x,y} & s_{x,z} \\ s_{y,x} & 1 & s_{y,z} \\ s_{z,x} & s_{z,y} & 1 \end{bmatrix} \tag{2}$$

where $s_{x,y} = s_{y,x} = avg\left[f^h_{y,x}, f^h_{x,y}, f^w_{y,x}, f^w_{x,y}\right]$ and $f^h_{x,y}$ denotes the proportion of husbands «x» getting married to women of «y» (script h stands for husband while w for wife). The matrix is symmetric and equal to 1 on the diagonals as similarity with own ethnicity is 1. I construct a similarity matrix as above for each enumeration area in the data-set. Based on the above matrix, S, the *Generalized Ethno-Linguistic Fractionalization* (GELF) indicator, as proposed in Bossart *et al* (2008), is as follows:

$$GELF = 1 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} s_{i,j} \tag{3}$$

where $i, j \varepsilon \{x, y, z\}$ and $s_{i,j}$ is an element of the matrix S.

It is important to note that GELF is particularly dependent on the variable(s) of association or similarity used to compute the matrix. In this paper, the similarity is based on cross-ethnic marriage which is computed from the couples data-set and does not necessarily reflect the overall heterogeneity within the community.

Another index of diversity computed is the *Entropy Index* (EI), also known as the Shannon Index, used for species to measure biodiversity. The ELF index gives bigger weight to higher proportion of ethnicities as it has a quadratic form while the EI gives a decreasing weight as the proportion of a specific ethnicity increases in a community. It takes a maximum value when all the ethnicities are equally represented in the

community. I compute a normalized EI, between [0,1], in each enumeration area:

$$EI = \frac{-\sum_{j=1}^{n} s_i ln(s_i) - \left[\frac{(n-1)}{2N}\right]}{ln(n)} \quad (4)$$

where s_i is the proportion of ethnic i in the community, n is the number of ethnicities in the community and N is the total number of individuals in the community.

The above indicators are computed on ethnicities in each enumeration area by exploiting data at the individual level. There are a total of 50 ethnicities in Cameroon, 67 in Ethiopia, 15 in Kenya and 9 in Malawi. The correlation between the three indicators I computed is 0.79 between ELF and GELF, 0.94 between ELF and EI and 0.75 between GELF and EI.

2.2 Empirical Strategy and Results

I use the most recent waves of DHS household surveys: Cameroon 2004, Ethiopia 2005, Kenya 2003 and Malawi 2004. They include both ethnicity and AIDS Indicator Survey (AIS). Due to spatial correlation on HIV prevalence these countries are broadly comparable. Summary statistics are provided in Table 1.

The DHS surveys collect information on population, health, HIV, and nutrition every five years in different countries. Together with the questionnaire, individuals are asked to give a blood spot for HIV test and the outcome is not communicated to the interviewees. In case individuals are interested, they are given a voucher for HIV test that can be consumed in the nearest hospital. The percentage of individuals who know their status before the DHS survey is very low. As such, DHS gives several advantages: it provides the most accurate estimate of HIV prevalence in the population; individuals are not aware of their status allowing for research on the determinants of HIV; individuals' HIV status is linked to their socio-economic characteristics and that of their partner.

Tab. 1: Descriptive Statistics

	HIV	Extr-mar. Sex	ELF	GELF	ENTROPY	Education	Poorest	Poorer	Middle	Richer	Richest	Migrant
Cameroon												
Man												
Mean	0.04	0.94	44.83	41.79	52.59	3.99	0.15	0.18	0.19	0.23	0.26	0.44
SD	0.19	2.09	28.98	29.60	28.39	1.99	0.35	0.38	0.39	0.42	0.44	0.50
Obs	5044	5036	5044	4917	4917	4518	5044	5044	5044	5044	5044	4917
Woman												
Mean	0.07	0.29	42.20	40.28	50.21	4.01	0.18	0.18	0.19	0.21	0.23	0.44
SD	0.25	0.85	29.46	29.92	29.07	1.94	0.39	0.38	0.40	0.41	0.42	0.50
Obs	5155	5137	5155	5041	5041	4130	5155	5155	5155	5155	5155	5041
Total												
Mean	0.05	0.61	43.50	41.03	51.38	4.00	0.17	0.18	0.19	0.22	0.25	0.44
SD	0.22	1.62	29.25	29.77	28.76	1.97	0.37	0.38	0.40	0.41	0.43	0.50
Obs	10199	10173	10199	9958	9958	8648	10199	10199	10199	10199	10199	9958
Ethiopia												
Man												
Mean	0.01	0.05	13.37	20.07	30.39	3.38	0.18	0.20	0.18	0.20	0.24	0.14
SD	0.09	0.26	20.69	24.86	32.50	1.78	0.39	0.40	0.38	0.40	0.43	0.34
Obs	5108	6026	6033	5980	5980	3599	6033	6033	6033	6033	6033	5980
Woman												
Mean	0.02	0.02	14.01	20.02	31.47	3.15	0.19	0.20	0.19	0.18	0.25	0.21
SD	0.13	0.22	21.46	24.81	32.86	1.77	0.39	0.40	0.39	0.39	0.43	0.41
Obs	5942	6793	6812	6750	6750	2734	6812	6812	6812	6812	6812	6750
Total												
Mean	0.01	0.04	13.71	20.04	30.96	3.29	0.18	0.20	0.18	0.19	0.25	0.18
SD	0.11	0.24	21.10	24.83	32.70	1.78	0.39	0.40	0.39	0.39	0.43	0.38
Obs	11050	12819	12845	12730	12730	6333	12845	12845	12845	12845	12845	12730
Kenya												
Man												
Mean	0.05	0.39	21.13	20.01	30.98	5.03	0.15	0.17	0.18	0.22	0.27	0.36
SD	0.21	0.95	26.33	26.65	32.45	2.39	0.36	0.38	0.39	0.42	0.45	0.48
Obs	2917	3573	3578	3480	3480	3278	3578	3578	3578	3578	3578	3480
Woman												
Mean	0.09	0.13	20.74	20.65	31.10	5.27	0.15	0.18	0.19	0.21	0.26	0.49
SD	0.28	0.37	26.20	26.82	32.45	2.37	0.36	0.38	0.39	0.41	0.44	0.50
Obs	3273	4030	4043	3955	3955	3406	4043	4043	4043	4043	4043	3955
Total												
Mean	0.07	0.25	20.92	20.35	31.04	5.15	0.15	0.18	0.18	0.22	0.27	0.43
SD	0.25	0.72	26.26	26.74	32.45	2.39	0.36	0.38	0.39	0.41	0.44	0.50
Obs	6190	7603	7621	7435	7435	6684	7621	7621	7621	7621	7621	7435
Malawi												
Man												
Mean	0.10	0.25	36.96	36.80	53.08	4.48	0.13	0.20	0.23	0.23	0.21	0.42
SD	0.30	0.87	26.33	25.47	28.60	2.29	0.33	0.40	0.42	0.42	0.41	0.49
Obs	2404	2401	2404	2375	2375	2150	2404	2404	2404	2404	2404	2375
Woman												
Mean	0.14	0.07	35.93	35.56	52.86	4.48	0.17	0.21	0.22	0.22	0.19	0.44
SD	0.35	0.27	25.69	25.55	28.43	2.30	0.38	0.40	0.41	0.41	0.39	0.50
Obs	2864	2864	2864	2828	2828	2190	2864	2864	2864	2864	2864	2828
Total												
Mean	0.12	0.15	36.40	36.13	52.96	4.48	0.15	0.20	0.23	0.22	0.20	0.43
SD	0.33	0.63	25.99	25.52	28.50	2.30	0.36	0.40	0.42	0.42	0.40	0.50
Obs	5268	5265	5268	5203	5203	4340	5268	5268	5268	5268	5268	5203

Both HIV status and extramarital sex (including premarital sex) vary across regions and countries. Men declare more extramarital sex than women but they are less likely to be HIV positive with respect to women. Cameroon is the most heterogeneous country in terms of ethnicity followed by Malawi, Kenya and Ethiopia. For ease of interpretation, all indicators of heterogeneity are used in percentage forms throughout

the paper. In terms of socio-economic characteristics, these countries are comparable except for education in Kenya, which is the highest. Further descriptive statistics can be found in Table 1.

Research on risky sexual behavior, or number of sexual partners, based on self-reported information suffers from under/over reporting. Their reliability depend on socio-economic characteristics of the respondent, the way the questionnaire was administered and last but not least quality of interviewers. Many studies have underlined how self-reported variables on sexual behavior result inconsistent when information is matched with biological features like sexually transmitted diseases (Gersovitz *et al.*, 1998; Glynn *et al.*, 2001).[8] For example, the number of extramarital sex, declared by both married and cohabiting individuals is inconsistent with their recent sexual activity in the last four weeks. I use data from monogamous couples where information on their sexual activity in the last four weeks and number of extramarital sex in the last twelve months is available both for the husband and wife. There are a total of 7,698 couples interviewed. In 1,046 cases, the declaration of sexual activity is discordant, *i.e.* husband declares to have been sexually active in the last four weeks while the woman not. Indeed, in monogamous couples, if one partner declares not to be active in the last four weeks while the other declares so, then the latter should declare at least one extramarital sex in the last twelve months. Out of 513 cases, where the wife was sexually active and the husband not, 99% of the women declared zero extramarital sex in the last twelve months. Extramarital sex is under-reported for men as well: only 17.5% of them declare at least one extramarital sex, even if they have been active in the last four weeks while their wife not. It is interesting to note that the discordance in recent sexual activity is, approximately, equally distributed among men and women (533 versus 513). This means that they were either negligent in reporting recent sexual activity leading to inconsistency or husband and wife under-report extramarital sex. One should correct for under/over reporting of extramarital sex with good instruments that predict reporting but not correlated with sexual behavior. It is however difficult to find variables with reliable exclusion restriction on reported extramarital sex. For this reason, I focus on the relationship between HIV prevalence and ethno-linguistic heterogeneity as HIV epidemics in Africa is known to be driven by heterosexual relations. In fact, recent studies on the subject complement self-reported sexual behaviors with other biomarkers of sexually transmitted diseases.

[8] Glynn *et al.* (2001) showed that 12% of women who reported being virgins were HIV positive and some had other sexually transmitted infections. This is labeled as "*social desirability bias*" where individuals tend to bias responses based on what is expected from the society.

The analysis of the chapter is conducted on HIV status of individuals, an objective measurement of risky sexual behavior. It is conditional on the assumption that the main channel of HIV transmission is heterosexual sex. In addition, exposure of men and women to HIV infection per sexual act differs biologically. This study includes controls for gender in the pooled analysis and in some cases I also conduct separate regressions by gender. As robustness checks, I further support the findings on the number of extramarital sex in the last 12 months.

The impact of Ethno-Linguistic Fractionalization on HIV Status

HIV prevalence rate ranges between 1.3% in Ethiopia and 12.3% in Malawi.[9] It is higher among women (1.7% in Ethiopia and 13.9% in Malawi) rather than men and in urban than in rural areas. I estimate the impact of ethnolinguistic heterogeneity on HIV status of individuals based the following equation:

$$HIV_{i,c} = \alpha + \beta H_c + \gamma X_{i,c} + \delta D_r + \varepsilon_{i,c} \quad (5)$$

where $HIV_{i,c}$ is HIV status of individual i in community c, H_c is ethno-linguistic or religious heterogeneity in the community, $X_{i,c}$ is a set of individual control variables in community c and D_r is a set of dummies at regional level.

Location in urban/rural areas, education and wealth might be, to some extent, choice variables for the individual or her family. Individuals migrate, due to their HIV status, in more heterogeneous areas. For-example, ethno-linguistic heterogeneity might give an opportunity to HIV positive individuals to keep confidentiality on their status and use strategic behavior with their partner in order not to infect him. It might also induce migration to seek medical care in cities which are likely to be heterogeneous. Notwithstanding, it implies that she is aware of her status, which is often not the case in the DHS. Hence, individuals are less likely to self-select in heterogeneous societies due to their HIV status. Even though migration might not cause a problem of reverse causality, it might bias the parameter of interest due to omitted variables. Migration and commercial areas are among the main drivers of HIV prevalence and at the same time they are likely to be in heterogeneous areas. I run the regressions on a sample of communities with baseline density and 70% of the people sampled in the community being residing for at least 10 years.[10] The restricted sample is, hopefully,

[9] All HIV prevalence rates in this paper are calculated by using the DHS sample weights

[10] I restrict the sample to enumeration areas where at least 10 people have been sampled. This allows a better interpretation of the indicators as they are less representative on small number of observations.

representative of areas which are less subject to inflow of migration like commercial areas. The restriction on the sample does not, however, rule out areas which are subject to seasonal migration. Nonetheless this does not affect the main conclusion of the paper.[11] I further control for migration where I define migrant a person who is residing in the place of residence for less than 10 years. The inclusion of regional fixed effects reduces the coefficient on ethno-linguistic heterogeneity. This is due to the fact that regional dummies capture ethnolinguistic heterogeneity at regional level. When not controlled for, they have an upward bias on the parameter of interest because positively correlated with ELF and HIV. Their inclusion implies that the coefficient on ELF is capturing heterogeneity within regions.

For illustrative purpose and comparison with Pongou (2011), I report the effect of ELF on HIV status and prevalence at enumeration area level in Table 2.[12] A one percent increase in ethno-linguistic heterogeneity increases the probability of being HIV positive by almost 0.0003. In relative terms, this corresponds to 0.5% increase. These coefficients are very similar to results found in Pongou (2011). Alternatively, it increases HIV prevalence rate at enumeration area level by .0366 (58.6%). I report results only on ELF for ease of comparison with Pongou (2011).

ELF is taken as a proxy for opportunities in networking (sexually) with individuals from other ethnicities. I test this mechanism on a dummy that represents a minority (or majority) ethnic group in a community. Minority (majority) dummy is more robust compared to an indicator computed at enumeration area level and dependent on different factors like number of observation and algebraic properties of the formula. The assumption behind is minority groups have more outside option than majority groups. For ease of interpretation, I exclude areas where two or more ethnicities are both minority and majority groups at the same time. In Table3 I show minority groups are likely to be HIV positive by 0.02 points while belonging to a majority group reduces it by .01.

[11] The DHS is a population based survey where randomly selected households are interviewed. As household members of seasonal migrants are less likely to move with the individual, it is likely that they do not affect the indicators of heterogeneity. These are often rural areas and likely to be homogenous. Thus, the restriction of the sample to non-migrant individuals will imply a downward bias and does not affect the main findings.

[12] As in Pongou (2011), only OLS coefficients are reported.

Tab. 2: Ethno-linguistic Heterogeneity on HIV status and HIV prevalence

OLS Regressions	Dependent Variables			
	HIV Status	HIV Prevalence	HIV Status	HIV Prevalence
ELF	0.000544***	0.0551***	0.000316***	0.0366***
	(7.92e-05)	(0.00749)	(8.92e-05)	(0.00788)
Religeous Fractionalization			-8.66e-05	-0.0133
			(0.000168)	(0.0151)
ETHNIC FE	YES	YES	YES	YES
REGIONAL FE	YES	YES	YES	YES
URBAN			YES	YES
SOCIO-ECONOMIC CONTROLS			YES	YES
Constant	0.0185	1.482	-0.0426	1.785
	(0.0237)	(1.218)	(0.0298)	(2.005)
Observations	30,139	34,070	30,139	34,070
R-squared	0.056	0.306	0.063	0.316

Standard errors are clustered at EA level. *** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variables are HIV status and HIV prevalence calculated at the enumeration area level. All regressions include regional dummies and ethnic fixed effects. Socio-economic controls include: wealth, education, religion and age. Although results from OLS are shown results hold with non linear models. In all regressions, sample weights are used and robust standard errors are clustered at the enumeration area level. Observations differ across the two samples due to attrition on HIV status at the individual level while HIV prevalence is calculated at the enumeration area (EA) level. In all regreassions sample weights are used and robust standard errors are clustered at enumeration area level.

Tab. 3: Belonging to a Minority/Majority group on HIV status

OLS regressions	Dependent Variable: HIV Status			
Belonging to the Majority Group	-0.0169***	-0.00868		
	-0.00519	-0.00536		
Belonging to the Minority Group			0.0211**	0.0206**
			-0.00838	-0.00839
Socio-Economic Controls		Yes		Yes
Constant	0.129***	0.0148	0.114***	0.00177
	-0.0374	-0.0409	-0.0368	-0.0403
Observations	22,248	22,248	22,248	22,248
R-squared	0.046	0.057	0.046	0.057

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: The dependent variable is HIV status of individuals and Belonging to a Majority (Minority) Group is a variable that takes the value 1 if the individual belongs to an ethnic group which is the most (least) sampled at the enumeration area level. All communities where only one ethinicity cannot be identified as the majority (minority) are excluded from the regression. Those are homogeneous areas as well as areas where all ethnicities are equally represented. The set of control variables are wealth, education, religion, age, religious fractionalization, ethnicity and regional fixed effects. In all regreassions sample weights are used and robust standard errors are clustered at enumeration area level.

Table 4 reports OLS regressions by disentangling the effect of gender from ethno-linguistic heterogeneity. Gender difference on the role of heterogeneity on HIV is put in evidence in the table. All the indicators have a positive effect on HIV status of women with broadly similar magnitudes. Indeed, I find a negative or no effect of heterogeneity on HIV status of men.

The theoretical foundation of GELF imply it is the best measure of heterogeneity and coherent with the theory of sexual networking. On the other side, GELF is computed on the data-set of couples as similarity is measured in terms of cross ethnic marriages. This implies all ethnicities of individuals not married or cohabitant are not included in the computation. Still, they might represent an outside option for married individuals who want to engage in risky behavior. In the regression, being a woman increases the probability of being HIV positive by one percent on average in a completely homogeneous community. This effect is additionally increased by ethno-linguistic composition of the community by 0.6 percent. The coefficient on gender captures the biological discrepancy of HIV infection between men and women. Results in Column (1) of Table 4 are based on the ELF, while Column (2) reports the Generalized Index and Column (3) shows the Entropy index. I further extend my analysis on the role of women's HIV status in married couples.

In Table 5, the analysis is based on sub-samples of women and men with their marital status. In Column (1), I show results from the pooled sample; from Column (2) to Column (7) the regressions are based on individuals who reside in enumeration areas where the number of people sampled is superior to ten and the proportion of migrants is less than 30 percent. This restriction reduces bias due to omitted variables caused by commercial areas. The sample is further restricted by gender and marital status.[13] An interesting outcome of the table is the role of heterogeneity on HIV status of women and specially married or cohabiting women at the time of the survey. An interpretation of the parameters would suggest that HIV positive and married women are likely to live in heterogeneous societies. Same interpretation cannot be extended to non married or cohabiting women, as well as men independently from their marital status. In the next section, I will focus on discordant couples where either she or he are HIV positive

Tab. 4: The Effect of Ethno-linguistic Heterogeneity on HIV Status by Gender

Dependent Variable: HIV STATUS			
OLS	(1)	(2)	(3)
ELF	-0.000177*		
	(9.61e-05)		
GELF		-0.000128	
		(7.80e-05)	
EI			-0.0158**
			(0.00682)
(ELF/GELF/EI) x Female	0.000650***	0.000440***	0.0427***
	(0.000143)	(0.000117)	(0.00998)
Female	0.00866***	0.0109***	0.00853***
	(0.00276)	(0.00280)	(0.00283)
Religious fract.	-0.000176	-0.000165	-0.000167
	(0.000207)	(0.000210)	(0.000211)
Muslim	0.00130	0.00179	0.00199
	(0.00322)	(0.00326)	(0.00326)
Urban	0.0153**	0.0181**	0.0185**
	(0.00777)	(0.00841)	(0.00814)
Age	0.000763***	0.000767***	0.000759***
	(0.000119)	(0.000120)	(0.000120)
Constant	0.00195	0.00218	0.00239
	(0.0187)	(0.0238)	(0.0193)
Obs.	14,937	14,848	14,848
R-squared	0.043	0.042	0.043

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: All regressions include 34 regional fixed effects, dummies for wealth and education. Sample is restricted to those areas where the number of people sampled in the enumeration area is superior to 10 and the proportion of migrants in the enumeration area is less than 30%. Results are stronger if we relax this restriction. In all regressions, samples weights are used and robust standard errors clustered at the enumeration area level.

in order to shed light on the role of heterogeneity on women.

Even though HIV status is a reliable information compared to self-reported variables, it is based on voluntary testing. Individuals can refuse to be tested. On average response rate is above 70% and it is higher for women rather than for men. It is also higher in rural areas. Descriptive statistics based on the analysis of mean differences of response rate among women suggest that those who refused are more likely to be HIV positive (Juhn *et al.*, 2009).[14] This tends to underestimate HIV prevalence. Non

[13] The sub-samples by marital status might not sum to the total number of men and women as I exclude those who are divorced, widowed or married but not living together at the time of the survey.

[14] Juhn *et al.* (2009) find that those who refuse are more likely to be educated, less likely to live in

Tab. 5: HIV Status and Different indicators of Ethno-linguistic Heterogeneity

Dep Variable	HIV Status						
	Total	Women			Men		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Probit		Total	Married	Not Married	Total	Married	Not Married
ELF	0.00348**	0.00568***	0.00553***	0.00405	-0.00186	-0.00262	0.00155
	(0.00154)	(0.00165)	(0.00194)	(0.00327)	(0.00233)	(0.00297)	(0.00383)
Obs.	14,824	7,891	5,204	2,201	6,446	3,455	1,932
GELF	0.00198	0.00338**	0.00474**	-0.00115	-0.000382	-0.00168	0.00354
	(0.00137)	(0.00156)	(0.00196)	(0.00256)	(0.00197)	(0.00254)	(0.00312)
Obs.	14,739	7,844	5,185	2,185	6,416	3,454	1,909
EI	0.00175	0.00298**	0.00409***	-0.00257	-0.00116	-0.000748	-0.00208
	(0.00123)	(0.00133)	(0.00149)	(0.00262)	(0.00194)	(0.00237)	(0.00336)
Obs.	14,739	7,844	5,185	2,185	6,416	3,454	1,909

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: All regressions include regional fixed effects, religious heterogeneity, religion, wealth, education, age and if rural or urban area. Results hold even if I include ethnic fixed effects and other socioeconomic characteristics. From column 2, the sample is restricted to those areas where the number of people sampled in an enumeration area is superior to 10 and 70% of the people sampled in a give community are residents for more than 10 years. Column 2 is based on the sample of all eligible women while column 3 and 4 are respectively Married and Unmarried women. I replicate similar analysis on men from column 5 to 7.

response will be explicitly addressed in the last section of the chapter.

2.3 Discordant Couples and Ethno-Linguistic Fractionalization

For better understanding the difference between serostatus of men and women and if infection occurred due to own behavior or that of partner's, I focus on couples which are HIV concordant or discordant. HIV concordant couples are married or cohabiting partners for at least five years and their serostatus coincides-HIV positive or HIV negative. If their serostatus does not coincide, they can be an HIV discordant couple where either the woman or the man is HIV positive. To this aim, I exploit the date-set of couples where HIV status is discordant. Out of 7,171 couples interviewed with information on their serostatus, 5.5% of them are in a discordant couple. 424 women and 411 men are HIV positive with married or cohabiting marital status.

Discordance among couples might occur due to risky sexual behavior with extramarital partners before or after marriage. It also depends if the couple is polygamous i.e. infection of a man might have occurred from own second partner and still be in a discordant couple. Approximately, all HIV positive married women and men are equally distributed between concordant and discordant couples. By restricting the sample to couples married for a determined period of time, i.e. ten years if data allows or five if else, HIV discordant cases are good proxies for effective extramarital sex[15]. By effective extramarital, I refer to HIV infection that occurred due to sexual encounter outside wedlock, given the partner is HIV negative.

Table 6 shows results on discordant couples. The proportion of men and women who are in a discordant status are comparable. I first consider the impact of ethno-linguistic heterogeneity on the probability of being in a discordant couple. Ethno-linguistic heterogeneity is positively associated with the probability of being in a discordant couple and it approximately measures effective risky behavior of both the husband and wife given the couple is married for a certain period of time. The effect on discordance is then split by wife's or husband's serostatus on a sample of couples who are married or cohabiting for at least five years. Results suggest a positive effect of ethno-linguistic heterogeneity on the probability of being in a discordant couple. The effect is stronger for discordant couples with the wife being HIV positive as shown in Column (3) and (4) . The excluded categories are concordant couples and discordant couples when the husband is HIV positive. The analysis on men is presented in Column (5) and (6).

rural areas and more likely to be in the wealthiest quintile. This is a similar pattern which separates HIV positive and HIV negative women.

[15] In some cases, I restrict the sample to couples married for 5 years in order to increase sample size.

Tab. 6: HIV Status and Discordance Among Couples

Probit	Dependent Varaibles: Discordance among couples					
	Concordant Couples		Discordant: She Positive		Discordant: He Positive	
Column	(1)	(2)	(3)	(4)	(5)	(6)
ELF	-0.00764***	-0.00320**	0.0116***	0.00590*	0.00158	-0.00326
	(0.00109)	(0.00139)	(0.00261)	(0.00333)	(0.00302)	(0.00388)
Observations	6,873	6,468	2,859	1,927	2,859	2,204
EI	-0.00670***	-0.00224*	0.00899***	0.00287	0.000422	-0.00395
	(0.00106)	(0.00124)	(0.00218)	(0.00242)	(0.00235)	(0.00332)
Observations	7,171	6,766	2,889	1,944	2,889	2,224
GELF	-0.00558***	-0.00219	0.00814***	0.00218	0.00129	-0.00207
	(0.00112)	(0.00135)	(0.00240)	(0.00333)	(0.00269)	(0.00376)
Observations	7,171	6,766	2,889	1,944	2,889	2,224
Region FE		YES		YES		YES
Wealth		YES		YES		YES
Education		YES		YES		YES
Muslim		YES		YES		YES
URBAN		YES		YES		YES
Age		YES		YES		YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: All regressions include regional fixed effects, religious heterogeneity, religion, wealth, education, age and if rural or urban area.The dependent variable is binary and takes the value one in Column (1) and (2) if the couple is concordant, positive or negative. From Column (3)-(4) the sample is restricted to couples with at least 5 years of marital duration and takes the value one whenever she is positive and he is negative. Finally in column (5) and (6) it takes the value one whenever he is poitive and she negative. In all regressions, sample weights are used and robust standard errors are clustered at the enumeration area level.

Table 6 shows HIV positive women are more likely to be in a discordant couple, with infection more likely to have occurred after marriage and in areas with high values of ethno-linguistic heterogeneity. Even though women are biologically more exposed to HIV than men, the restriction of the sample to a minimum of five year marriage allows to infer that their infection occurred with extramarital relation.[16] Column (3) and (4) highlight the role of gender and it distinguishes the paper from results presented previously in the literature; it shows that women are more likely to be infected with HIV from out-of-wedlock sexual relation in heterogeneous areas, while the same does not hold for men's infection.

2.4 Sensitivity Analysis

The Effect of ELF on Risky Sexual Behavior

There are numerous factors that influence individual's risky sexual behavior: general individual level characteristics (age, education, wealth, marital status, health), social characteristics (communities' ethnic or religious identities), knowledge and exposure to HIV/AIDS and cost of HIV infection and prevention. In the sample, 17% of individuals declared at least one extramarital sex in the last twelve months with relevant heterogeneity among countries. Out of the total number of men sampled, 25% declared at least one versus 11% of women, showing that women reported less risky sexual behavior than men.

I test if results found on HIV test are reflected on declared extramarital sex. I define extramarital sex as sex with non marital or cohabiting partner. The mechanism through which ethno-linguistic heterogeneity impacts HIV status is the same through which it impacts extramarital sex. It is believed that the main route of HIV transmission in sub-Saharan Africa is heterosexual sex. The impact of the indicators on risky sexual behavior is evaluated with the following reduced form:

$$s_{i,c} = \alpha + \beta H_c + \gamma X_{i,c} + \delta D_r + \varepsilon_{i,c} \quad (6)$$

where $s_{i,c}$ is the number of *declared* extramarital relations of individual i in community c. Following earlier definition, H_c is ethno-linguistic or religious heterogeneity in the

[16] The conclusion is based on the assumption that the main channel of HIV infection is heterosexual sex.

community, $X_{i,c}$ is a set of individual control variables in community c and D_r is a set of dummies at regional level.

Table 7 reports results based on Poisson regression of *declared* extramarital sex on heterogeneity and other control variables. As in the above analysis, there is a positive and significant impact of ethno-linguistic heterogeneity on the number of *declared* extramarital relation.[17] Married women report higher number of extramarital sex in more heterogeneous societies. Similarly, married men also report higher number of sexual partners in heterogeneous societies, but the magnitude is lower. For married women, the table suggests a 0.0135 points increase in the difference in logs of expected number of extramarital partners in the last 12 months when heterogeneity increases by one percentage points keeping the other variables constant. This implies that married women, to some extent married men as well, tend to engage in risky sexual behavior in heterogeneous societies by confirming the above analysis.

Public Good Provision

Ethno-linguistic heterogeneity and public goods provision has been discussed in the literature for its role on conflicts and heterogeneous preferences, as well as tastes (Alesina *et al.*, 1999; Alesina *et al.,* 2000; Alesina *et al.*, 2003 ; Alesina *et al.*, 2005). In these studies, a possible mechanism through which ethno-linguistic heterogeneity affects HIV status might be due to public policies or public good provision at community or regional level. In light of this, regional and rural/urban dummies should capture any difference in the provision of public policies related to HIV/AIDS epidemics. Those dummies capture any public policy that is specific to the region. Indeed a common characteristic of African countries is the role played by policy makers in enhancing regional and ethnic favoritism in the allocation of governmental funds or public good provision, as emphasized in Lierberman (2007). I argue that the dummies capture potential bias due to ethnic favoritism in addressing government funds against HIV at regional level. Moreover, results do not change when replicated with ethnic fixed effects.

[17] As in the above regressions, the database includes all eligible men and women who are resident in those enumeration areas with at least 10 individuals sampled and with a baseline density. I also restricted the sample to those areas where 80% of the individuals are residents for at least 10 years.

Tab. 7: Extramarital Sex and Ethno-linguistic Heterogeneity

	Dependent Variable: Extramarital Sex						
	Total	Woman			Men		
Poisson	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Total	Married	Not Married	Total	Married	Not Married
ELF	0.00576*	0.0138***	0.0135***	0.00681	0.00297	0.00680*	0.000687
	(0.00304)	(0.00426)	(0.00504)	(0.00485)	(0.00366)	(0.00412)	(0.00430)
Religious Fractionalisation		-0.00780	-0.0219***	0.0165	-0.000361	0.00452	-0.00361
		(0.0107)	(0.00782)	(0.0139)	(0.00585)	(0.00589)	(0.00776)
Region FE	YES	YES	YES	YES	YES	YES	YES
Wealth	YES	YES	YES	YES	YES	YES	YES
Education	YES	YES	YES	YES	YES	YES	YES
Muslim	YES	YES	YES	YES	YES	YES	YES
URBAN	YES	YES	YES	YES	YES	YES	YES
Age	YES	YES	YES	YES	YES	YES	YES
Obs.	16,535	8,736	5,795	2,941	7,799	4,334	3,429

Robust standard errors are clustered at the enumeration area level. *** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is a count variable: the number of extramarital sex in the last 12 months. It includes premarital sex for individuals who are not yet married. In column II, we restricted the sample to eligible women while in 3 and 4 we distinguish between married and unmarried. The same type of analysis is extended to men in columns 5 ot 7. In all regressions, sample weights are used and robust standard errors are clustered at the enumeration area level.

Compliance in HIV Test

Response rate in the data-set varies across countries, urban/rural areas and gender. Compliance is approximately 4.4% in Cameroon, 14.4% in Ethiopia, 18.5% in Kenya and 23.9% in Malawi. Different levels of response rates might over/underestimate HIV prevalence and bias the parameter of interest if compliance is correlated with ethno-linguistic heterogeneity. In Figure1-3, I show the correlation between these variables. On the vertical axis of Figure 1 is the quintile distribution of HIV prevalence at enumeration area level while in Figure 2 and Figure3 is the quartile distribution of the ELF. They are all positively correlated implying a potential bias in the estimator.

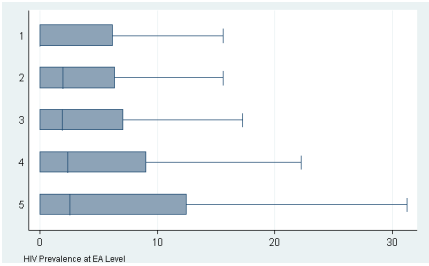


Fig. 1: HIV Prevalence by Compliance Quintiles.

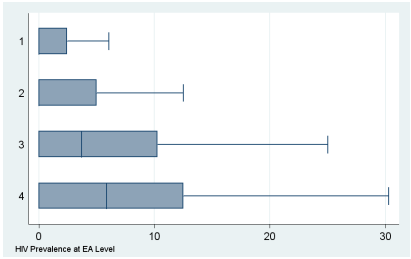


Fig. 2: HIV Prevalence by ELF Quantiles

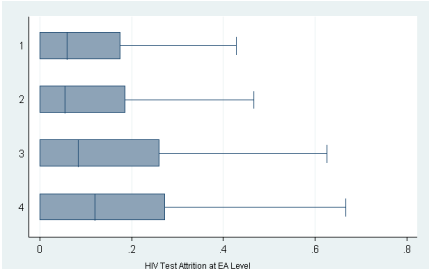


Fig. 3: HIV Compliance by ELF Quantiles

Table 8 presents mean differences between compliers and non. The ELF is statistically different among the two groups. Those who do not comply for an HIV test are more likely to be richer, educated and never married. They are also more likely to be migrants and to be living in urban areas, where on average ELF is higher. In terms of risky sexual behavior, non compliers tend to report low number of risky behavior and to have been tested for HIV [18]. The bias of the parameters is sever to the extent non compliance is large or if compliance predicts the probability of being HIV positive and, finally, to the extent compliance is correlated with ELF.

Tab. 8: Socio-Economic Difference between Compliers and Non Compliers

		Mean				
Variable	*Obs Compliers*	*Obs Not Compliers*	*Compliers*	*Non Compliers*	*Difference*	
Ethnic Fractionalisation	45240	7259	32.43	37.47	5.04	***
Female	45240	7259	0.47	0.48	0.01	
Extram. Sex in 12 months	45179	7221	0.28	0.20	-0.09	***
Migrant	45240	7259	0.35	0.38	0.03	***
Age at First Sex	45172	7226	16.30	16.66	0.36	
AIDS Test	27472	4340	0.13	0.16	0.03	***

(a) Mean Difference between Compliers and Non Compliers of HIV Test

		Mean				
Variable	*Obs Compliers*	*Obs (Not Tested)*	*Compliers*	*Non Compliers*	*Difference*	
Age	45240	7259	29.2177	29.171	-0.074	
POOREST	45240	7259	0.1701	0.1529	-0.022	***
POORER	45240	7259	0.1781	0.1274	-0.051	***
MIDDLE	45240	7259	0.197	0.1314	-0.063	***
RICHER	45240	7259	0.1969	0.1617	-0.03	***
RICHEST	45240	7259	0.258	0.4265	0.166	***
NO EDUCATION	45240	7259	0.261	0.2476	-0.023	***
PRIMARY	45240	7259	0.4206	0.4091	-0.006	
SECONDARY	45240	7259	0.2809	0.2718	-0.005	
HIGHER	45240	7259	0.0376	0.0715	0.034	***
URBAN	45240	7259	0.3146	0.4156	0.103	***
MARRIED	45240	7259	0.5935	0.5838	-0.019	**
NEVER MARRIED	45240	7259	0.3301	0.3422	0.018	**
FORMERLY MARRIED	45240	7259	0.0765	0.074	-0.003	

(b) Mean Difference between Compliers and Non Compliers of HIV Test

To measure the severity of the bias, I replicate the above results on sub-samples based on areas with different levels of compliance: all compliant areas and areas above and below the median compliance rate. Table 9 shows the sign and significance of the

[18] There are studies stating that HIV prevalence, refusal and prior testing rates are highly correlated. Socio-economic characteristics which predict HIV status are more likely to predict non compliance. Moreover, Eaton *et al.* (2009) demonstrated that, in Malawi, HIV positive individuals who know their status are almost five times more likely to refuse testing than HIV negative individuals.

parameters in all the three sub-samples. In the first two columns, I report results on the overall sample for comparison. Column 3 and 4 restrict the sample on those enumeration areas where response rate is full. The magnitude of the parameter decreases, but the sign and significance persist. In Column 5 and 6, I present results on the sample of individuals who live in enumeration areas with compliance is above the median, while Column 7 and 8 present results for areas where compliance is below the median. The main difference between Column 4, 6 and 8 is the increasing number of compliance. The magnitude of the coefficient increases with the increasing number of compliance. They decrease when adding controls that are likely to predict compliance and HIV status like wealth, education and migration, as shown in Table 8. Compliance does not drive the main conclusion of our result. Indeed the effect of ethno-linguistic heterogeneity is not totally driven by sample selection because the coefficient on ELF persists in all the regressions, as shown in Table 9. Moreover, compliance is more likely to occur in areas where HIV prevalence is low, implying sample selection of homogeneous areas and downward bias in HIV prevalence. These areas are also likely to be homogeneous in ethno-linguistic heterogeneity. Comparing the magnitude of the coefficients on ethno-linguistic heterogeneity in Column 2, 4 and 6 results suggest the parameters increase with the degree of non-compliance in the sample.

Migration and Commercial Areas

The main conclusion of the paper might be driven by inflow of migration: presence of migrants in an area affect ethno-linguistic heterogeneity and at the sometime migrants are likely to practice high risk sexual activities compared to non migrants. In all regressions, we control for migration where I define migrant an individual who has not been living in the place of interview for more than ten years. Ten years is the minimum number to identify those who have contracted the virus in their actual place of residence.

Unfortunately, data do not allow to control for seasonal migration but being a phenomenon of rural areas, it does not affect the main conclusion of the paper.

Alongside the same reasoning, commercial or trading areas might bias the parameters if correlated with migration and transactional sex at the same time. To this purpose the sample is restricted, throughout the paper, to areas with 80% of the people interviewed being residents for at least 10 years. This should allow to rule out areas affected by in-or-out flow of migrants and commercial areas.

Tab. 9: The Effect of ELF on HIV status with different levels of compliance

Dependent Variable: HIV Status								
	Total Sample		All Compliers		% Complier above Median		% Complier below Median	
LPM	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ELF	0.000584***	0.000321***	0.000441***	0.000257**	0.000736***	0.000423***	0.000825***	0.000538***
	(7.49e-05)	(7.79e-05)	(0.000106)	(0.000108)	(0.000144)	(0.000145)	(0.000187)	(0.000202)
Regional FE	YES	YES	YES	YES	YES	YES	YES	YES
Female		0.0243***		0.0237***		0.0289***		0.0175**
		(0.00370)		(0.00511)		(0.00680)		(0.00858)
Richest		-0.0391***		-0.0432***		-0.0292***		-0.0458***
		(0.00715)		(0.0125)		(0.0107)		(0.0166)
Richer		-0.0337***		-0.0397***		-0.0261**		-0.0260
		(0.00751)		(0.0128)		(0.0110)		(0.0171)
Middle		-0.0208***		-0.0219*		-0.0212**		-0.00723
		(0.00668)		(0.0113)		(0.0102)		(0.0154)
Poorer		-0.0115*		-0.0104		-0.0127		-0.00436
		(0.00616)		(0.0110)		(0.00868)		(0.0148)
Primary		0.0170***		0.0195***		0.0157*		0.00961
		(0.00454)		(0.00510)		(0.00885)		(0.0120)
Secondary		0.0105*		0.00384		0.0194		0.00949
		(0.00602)		(0.00702)		(0.0123)		(0.0146)
Higher		-0.0131		-0.000145		-0.0172		-0.0209
		(0.0104)		(0.0208)		(0.0177)		(0.0184)
Migrant		0.0156***		0.0127**		0.0172***		0.0214***
		(0.00358)		(0.00539)		(0.00613)		(0.00750)
Urban		0.00978*		0.00789		0.0148		0.00885
		(0.00589)		(0.00856)		(0.0107)		(0.0152)
Married		-0.0708***		-0.0460***		-0.0717***		-0.118***
		(0.00745)		(0.0106)		(0.0119)		(0.0190)
Never Married		-0.109***		-0.0820***		-0.105***		-0.171***
		(0.00792)		(0.0116)		(0.0124)		(0.0192)
Constant	0.0411*	0.104***	-0.000303	0.0718***	0	-0.0413***	0.0273	0.139***
	(0.0238)	(0.0253)	(0.000425)	(0.0179)	(3.93e-07)	(0.0119)	(0.0273)	(0.0366)
Observations	30,139	30,139	13,133	13,133	10,366	10,366	6,640	6,640
R-squared	0.041	0.065	0.039	0.059	0.043	0.067	0.045	0.082

Robust standard errors are clustered at enumeration area level. *** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is the probability of being HIV positive. In the first column we represent results for the total sample, with and without controls, for ease of comparison. In column 3 and 4, we present results for areas where the percentage of compliers were 100% at enumeration area level. For areas where non response was above zero, we split the sample into 2: above or below the median value of non response rate. The median value was 15.8%. In Column 5 and 6 we present results where response rate was above 15.8 while in column 7 and 8 is the reverse.

I further show to what extent the results on ELF can be attributed to migration. I analyze the proportion of migrants at enumeration area in Table 10. The table shows how the parameter of interest changes with different levels of migration. Migration induces an upward bias on the parameter of interest. This holds in Column 2 and 3 with different levels of migration: years of residence below 10 and 5. Column 4 shows the coefficient on permanent residents compared to individuals not born in the place of interview. The impact of ELF on HIV status is robust to a set of controls and the indicators of migration (Column 5, 6 and 7).

2.5 Conclusions

This chapter used the most recent nationally representative surveys from four sub-Saharan countries, spatially comparable, to analyze the effect of ethno-linguistic fractionalization on HIV epidemics. The paper emphasizes the role gender plays when engaging in risky sexual behavior. Women's behavior tend to be more elastic to ethno-linguistic heterogeneity than men's. The hypothesis is tested on *declared* extramarital sex of women, their HIV status and the probability of being in a discordant couple where she is HIV positive. The study puts forward an original finding by emphasizing two channels which underline the role social environment plays on women's behavior and their HIV status. This role can be attributed to the probability of being detected from deviating behaviors, as emphasized in Pongou (2011), and gender discrepancy in the cost associated to social sanctions when engaging in extramarital sex.

The study further disentangles the effect of ethno-linguistic heterogeneity on own status that is attributable to own risky sexual behavior from that of the husband's risky behavior. Based on a sample of HIV positive individuals who are in a discordant couple, I test ethno-linguistic heterogeneity on the probability of being in a discordant couple and show that the indicators have a positive impact on women's HIV positive status with non infected husbands and infection occurred after marriage. This confirms that women's behavior tend to differ from men's when they internalize social costs due to deviating behaviors.

The paper contributes in underlining community level characteristics that shape individual's risky behavior that are necessary when addressing specific prevention policies. Mostly, the analysis contributes in better understanding the gender difference that exist when engaging in extramarital sex and the consequence on HIV prevalence. The paper sheds light on the different mechanisms that shape the patterns of HIV prevalence across regions, sex and socio-economic characteristics.

Tab. 10: The Effect of ELF on HIV status with different levels of Migration

OLS Regression	Dependent Variable: HIV Status						
ELF	*0.000544****	*0.000359****	*0.000377****	*0.000391****	*0.000213***	*0.000225***	*0.000252****
	(7.92e-05)	*(8.12e-05)*	*(8.06e-05)*	*(8.21e-05)*	*(8.80e-05)*	*(8.86e-05)*	*(8.95e-05)*
Migration Rate (less then 10 years)		0.000571***			0.000490***		
		(0.000116)			(0.000117)		
Migration Rate (less then 5 years)			0.000659***			0.000584***	
			(0.000142)			(0.000143)	
Always Resident Rate				-0.000479***			-0.000321***
				(0.000114)			(0.000112)
Socio-Economic Controls							
Constant	0.0185	0.00274	0.00788	0.0424*	-0.0503*	-0.0462	-0.0206
	(0.0237)	(0.0238)	(0.0242)	(0.0240)	(0.0295)	(0.0299)	(0.0305)
Observations	30,139	30,139	30,139	30,139	30,139	30,139	30,139
R-squared	0.056	0.057	0.057	0.057	0.064	0.064	0.064

Robust standard errors are clustered at the enumeration area level. *** p<0.01, ** p<0.05, * p<0.1

Note: The dependent variable is HIV status of individuals while Migration Rate is the percentage of people living less than 10 (or 5) years at the enumeration area level. Always Resident Rate is the percentage of people who have been living in their place of residence at the enumeration area level. The set of control variables are wealth, education, religion, age, religious fractionalization and regional fixed effects. In all regreassions sample weights are used and robust standard errors are clustered at enumeration area level.

3 Free Access to HAART and Pregnancy Response among HIV Patients. A Case Study from Cameroon

The emergence of Highly Active Anti-Retro-Viral Therapy (HAART) has allowed for sensible improvement in quality of life for people living with HIV/AIDS. In the past, with the Bamako Initiative-1987, many developing countries charged patients fees for health care delivery with the aim of cost recovering and health-care funding. Not by chance, in many Sub-Saharan African countries, more than half of the total health expenditure is borne by households. This persistence of user fees for healthcare, especially for HAART, decreased adherence and treatment effectiveness. The decrease in the price of treatment coupled with increasing donor support for HIV/AIDS programs has induced an incredible increase in access to treatment.

In Cameroon, until May 2007, HIV positive patients used to pay for treatment, laboratory tests and physicians' consultations. Since 2007, the government made HAART treatment free of charge for all HIV positive patients based on the WHO guidelines. By June 2008, 50000 patients i.e. 58% of the estimated total HIV-positive patients were benefiting from ART.[19] The policy has given many patients in Cameroon longer life expectancy and better health outcomes in a very short period (Wools-Kaloustian *et al.*, 2006). The benefits of HIV/AIDS treatment go beyond the direct health outcomes and impact socio-economic status of indivduals and ultimately micro-macroeconomic determinants of developing countries. Different dimensions of these treatment benefits have been object of study in the past 10-15 years.

While the direct benefits of treatment are large and known, a concern arises in balancing these benefits with the risk of HIV incidence. Health benefits of treatment are likely to influence individuals' perception on the epidemics and alter their sexual behavior. The change in risky sexual behavior can be due to reproductive choices or utility from non protected sex both among infected and non infected. The overall and net effect of treatment depends on several factors.

Firstly, the HAART therapy has a direct effect on life expectancy (Marins *et al.*, 2003; Goldie *et al.*, 2006). Not only individuals will be infectious for a longer period of time but they might also modify their reproductive choice for various reasons. The psychological cost of learning own HIV infection is perceived as reduced life expectancy be-

[19] Ministère de la Santé Publique and Comité National de Lutte contre le VIH/Sida du Cameroun. Vers l'accès universel au traitement et à la prise en charge du VIH/Sida chez les adultes et les enfants au Cameroun [in French]. Yaounde: Ministère de la Santé Publique and Comité National de Lutte contre le VIH/Sida du Cameroun; 2008.

cause there is uncertainty in the efficacy and availability of HAART treatment. Once individuals are on treatment, their choices on reproductive behavior and intended pregnancy depends on perceived life expectancy which they update by observing objective measures of their present health outcomes. There is no general consensus on the sign and magnitude of the theoretical and empirical prediction of adult mortality on fertility.

Secondly, with HAART individuals gain their full functional capacity by becoming less susceptible to other opportunistic diseases and attractive in the matching market (Koenig *et al.*, 2004; Laurent *et al.*, 2002; Coetzee *et al.*, 2004). Several papers found a positive impact of treatment on risky sexual behavior: Lakdawalla *et al.* (2004) among HIV positive individuals in the US; Mechoulan (2007) among homosexuals in San Francisco; Goldstein *et al.* (2007) in rural Kenya and De Walque *et al.* (2010) in Mozambique. Goldman *et al.* (2006) used an IV strategy to show that the provision of treatment increased the number of new partners for HIV positive individuals. On the other side, improved health outcomes of individuals increase the likelyhood of pregnancy among HIV positive patients because HAART ameliorates fecundity.[20] Thus, health benefits can increase both risky sexual behavior and unintended pregnancy among people living with HIV/AIDS.

Third, HAART reduces viral load and infectivity of individuals (Porco *et al.* 2004; Castilla *et al.* 2005). Availability of HAART and lower transmission of HIV reduces the cost of HIV infection and individuals might increase risky sexual behavior and unintended pregnancies. On the other side, non protected sex among individuals on treatment induces viral mutation and drug resistance. Furthermore, altruistic behavior of individuals might increase protected sex (Bunnell *et al.*, 2006). The net effect of HAART and reduced cost of infection on risky sexual behavior and non intended pregnancy remains an empirical question.

Finally, people on HAART treatment are usually enrolled in a treatment program which comprises counseling on HIV and sexual behavior. Many studies have addressed the impact of educational campaign on HIV/AIDS with mixed or no results on unprotected sex (Duflo *et al.*, 2006 and Bertozzi *et al.,* 2006). Dupas (2006) finds a positive effect of age-specific educational campaign on inter-generational sex for school girls.

The above channels induce behavioral change in risky sexual behavior and intended-or unintended-pregnancies. The overall net outcome remains an empirical question. The

[20] HAART restores menstrual disorder and reduces vaginal infections.

study takes advantage of firsthand data collected on women enrolled in a treatment program at the Central Hospital of Yaounde in 2010. The sample consists of HIV positive women, with part of them on treatment. I use the staggered timing of patient's date of treatment initiation between 2003-2010 to estimate the association between treatment and intended -or unintended-pregnancy. I use as controls, women who are not yet on treatment at the time of the survey. Pregnancy is an objective measurement of risky sexual behavior and understanding the role of HAART on the two types of pregnancy has different policy implications. I first implement a Diff-in-Diff approach in a moving-window set-up and further enrich the analysis by using panel data analysis where I control for individual and time fixed effects. The approach allows for individual specific trends where I separate pre-existing trends from dynamic response to treatment. I also extend the analysis by including a differential trend between the pre and the post treatment period.

Results suggest that HAART increased intended pregnancy among HIV positive women by 15 percentage points on average. These increase is heterogeneous by year of treatment initiation: the coefficient is increasing in magnitude for individuals who start treatment in recent years with a big jump being characterized in 2007. This is in line with the literature in epidemiology, suggesting treatment increases pregnancy among women living with HIV/AIDS. Additionally, treatment effect is heterogeneous by the number of children born before treatment initiation. Women starting treatment with zero number of children before treatment initiation are the most affected by treatment. Given the total fertility rate for women living in Yaounde amounts to three, I show that the effect of treatment is such that it does not increase the total fertility rate. Initial CD-4 also matters for pregnancy response; women who start treatment with CD-4 above the average are most likely to be pregnant after treatment initiation. The CD4 count is an important indicator of disease progression among HIV-infected individuals.

Different analysis suggest that the above results are not driven solely by the direct health effect of treatment related to risky sexual behavior and non intended pregnancy. Indeed, I show that results hold if I control for individuals who start treatment for reasons related to menstrual disorder. I further show that coefficients on treatment are time dependent, suggesting free access had an important role in fertility decision; finally, I find that treatment is most effective among a sub-population of individuals, i.e. with zero number of children before treatment initiation. These results provide evidence in support of a need for a broader interpretation of the behavioral response

of pregnancy that is not solely driven by health benefits.

To my knowledge, up-to-date nobody has analyzed the impact of ART treatment on pregnancy proxied for sexual behavior in the economics literature. In epidemiology, Myer *et al.* (2010) found that ART doubles the chances of becoming pregnant among HIV-infected women in sub-Saharan Africa. In their analysis, one third of women who initiated ART experienced pregnancy within four years after their initiation. In the same line, other papers have underlined that HAART restores fertility among HIV patients with fertility desire changing over time (see for example Blair *et al.* (2004) and Massad *et al.* (2004)). Among the few longitudinal studies on fertility preferences, Homsy *et al.* (2009) report that pregnancy significantly increased over follow-up. However, even though fertility intentions increased over time, they were much lower compared to the pregnancies encountered among couples. The behavioral mechanisms that underlie this association have not been further investigated by previous studies for lack of information.

The analysis proceeds as follows: data are presented in Section 3.1 and empirical strategy in Section 3.2. In Section 3.3, I present the results while in Section 3.4 and 3.4 heterogeneity of parameters are tested by sub-population. I discuss limitations of the results in Section 3.5 and concluding remarks are laid in Section3.6.

3.1 Data

The Day-Hospital in the Central Hospital of Yaounde has been active since 1998. It is a public structure, one of the biggest and the first to provide HIV treatment on a large scale. It offers several services among which: Voluntary Counseling and Testing, Sexually Transmitted Diseases Counseling, Dermatitis, Psychological Support, Social Assistance and General Counseling for people living with HIV. Once infected, patients can live for several years, with their health declining, until they reach the acquired immune deficiency syndrome status (AIDS). The virus's degree of advancement is measured by CD4 counts and once patients reach AIDS status they are put on treatment. [21] Else, their survival is very low due to opportunistic diseases. Patients monitor their CD4 count and viral load every 6 months until they become eligible. A Committee of medical doctors and personnel from the hospital decide on patient's eligibility. Once on treatment they renew their medical prescription every three months for HAART. The number of individuals who are detected HIV positive and put on

[21] The threshold of CD4 for treatment initiation, as suggested by the WHO, is a CD4 count of less than 350/mm3 for 2010 onwards. Previously the threshold was 200/mm3.

treatment have increased over years. Patients differ in terms of HIV detection and treatment initiation. About 400 HIV positive women on treatment and 135, not yet on treatment by 2010, were interviewed.[22] Most of the time, individuals get tested for HIV due to pregnancy and Program on Mother To Child Transmission (PMTCT); however, enrollment to treatment involves other reasons among which, dermatitis, TBC, amenorrhea, diarrhea or fever. Although data is not available for all, more detail is provided in Table 11. Shortage of personnel in the hospital implies that medical records are not fully compiled. Information for a total of 228 patients was collected from medical records.

Tab. 11: Medical Record on date of Treatment Initiation

AMENORHE	14	6.14%
ANEMIE	4	1.75%
DERMATITIS	86	37.72%
DIARHEE	16	7.02%
FEVER	27	11.84%
PREGNANCY	6	2.63%
HEPATITE B	2	0.88%
TBC/COUGH	20	8.77%
TRANSFERED	4	1.75%
WEIGHT LOSS	2	0.88%
NO FLAG	47	20.61%
Total	**228**	

I collected information on women aged between 15-49 years on self-reported knowledge about HIV/AIDS, as well as their sexual behavior, birth history and socio-economic indicators. Sexual behavior and pregnancy history was collected confidentially within the hospital. The main variable of interest childbearing where I distinguish between desired and non desired pregnancy. I retrospectively identified each pregnancy with birth recalling and constructed a panel which varies between 2003-2010. The conventional measure of intended/unintended pregnancy from the standard DHS surveys is used for ease of interpretation. Each questionnaire is matched then with objective information from their medical record where date of HIV test, treatment initiation, CD4 count and reasons for treatment and HIV detection was recorded.

Given the main variable of interest is birth history of women, a major drawback with recalling might be miss-reporting if women forget the birth of a child who died. One

[22] Very few patients had been on treatment for more than five years while a good percentage started in 2007.

can argue that women recall better the birth of a child after treatment because more recent and in which case biases the results upwards. However, information is collected on a restricted window and over the last seven years so it is less likely miss-recalling biases birth history of women. Moreover, women who discovered to be HIV positive are less likely to miss-recall sexual behavior or child birth because it might have affected partner's infection. Another variable of interest is intended birth history which is collected retrospectively, a potential drawback arises from measurement error. It is mainly based on ex-post rationalization of individual's desire for fertility and might be limited in its utility because it is based on the assumption all women have fully formed intentions at the time of conception. Bias is likely to arise from several factors like gender of the child, education of the mother, marital status, partner's intentions and etc.[23]

Almost all women in the data are aware of PMTCT and, therefore, the possibility for HIV positive women to give birth to a healthy child. There is no statistical difference on knowledge about all risks of vertical transmission between women who are on treatment and not. Indeed, HAART reduces child mortality among both groups of HIV positive women. PMTCT puts all pregnant women on treatment some months before and after childbearing. As such, they can avoid vertical transmission from the mother to the child while pregnant and during breastfeeding. When women enroll for PMTCT, they might, or not, already have reached AIDS status. If so, she will keep on with HAART treatment by renewing every three months her medical prescription. Else, she monitors her CD4 until eligible for treatment. The In Table 12, I provide descriptive statistics by year of treatment initiation, age, education and pregnancy pre and post treatment.

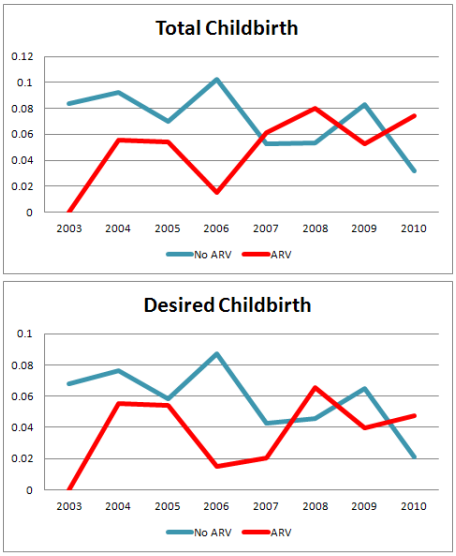
The average number children born after treatment increases over time as shown in Figure 4. Before 2007, women on treatment had lower number of births per woman as compared to HIV positive women not yet on treatment. However, after the policy there is a catch-up between intended and unintended births.

[23] A more complete analysis on measures of fertility preferences can be found in Pritchett (1994).

Tab. 12: Descriptive Statistics by Treatment Year

Year of Treatment	Age		Education		Fertility		Children before Treatment		Children after Treatment	
	Mean	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean	Obs
2000	**37.00**	**2**	**8.00**	**2**	**2.50**	**2**	**2.00**	**2**	**0.50**	**2**
	1.41	*74*	*2.83*	*16*	*0.71*	*5*	*1.41*	*4*	*0.71*	*1*
2001	**37.00**	**1**	**14.00**	**1**	**3.00**	**1**	**0.00**	**1**	**2.00**	**1**
	.	*37*	*.*	*14*	*.*	*3*	*.*	*0*	*.*	*2*
2002	**35.00**	**2**	**12.00**	**2**	**1.50**	**2**	**1.00**	**2**	**1.00**	**2**
	7.07	*70*	*1.41*	*24*	*0.71*	*3*	*0.00*	*2*	*1.41*	*2*
2003	**36.60**	**5**	**11.80**	**5**	**2.40**	**5**	**2.25**	**4**	**0.40**	**5**
	2.61	*183*	*1.64*	*59*	*2.30*	*12*	*0.96*	*9*	*0.89*	*2*
2004	**34.25**	**8**	**10.88**	**8**	**2.75**	**8**	**2.29**	**7**	**1.13**	**8**
	3.11	*274*	*3.23*	*87*	*1.91*	*22*	*1.60*	*16*	*0.99*	*9*
2005	**33.89**	**19**	**10.84**	**19**	**2.05**	**19**	**1.78**	**18**	**0.37**	**19**
	3.57	*644*	*3.06*	*206*	*1.27*	*39*	*1.17*	*32*	*0.76*	*7*
2006	**33.30**	**30**	**11.03**	**29**	**1.90**	**30**	**1.72**	**25**	**0.53**	**30**
	4.47	*999*	*3.44*	*320*	*1.47*	*57*	*1.31*	*43*	*0.68*	*16*
2007	**32.58**	**31**	**11.23**	**30**	**1.42**	**31**	**1.64**	**22**	**0.35**	**31**
	3.98	*1010*	*3.70*	*337*	*1.15*	*44*	*0.85*	*36*	*0.55*	*11*
2008	**32.08**	**39**	**9.84**	**37**	**1.90**	**39**	**2.00**	**32**	**0.36**	**39**
	4.91	*1251*	*3.06*	*364*	*1.47*	*74*	*1.19*	*64*	*0.63*	*14*
2009	**32.46**	**89**	**10.49**	**85**	**1.53**	**89**	**1.83**	**70**	**0.13**	**89**
	4.52	*2889*	*2.95*	*892*	*1.37*	*136*	*1.23*	*128*	*0.38*	*12*
2010	**31.65**	**69**	**11.35**	**66**	**1.81**	**69**	**2.19**	**57**	**0.00**	**69**
	5.34	*2184*	*3.42*	*749*	*1.54*	*125*	*1.42*	*125*	*0.00*	*0*
Total	**32.59**	**295**	**10.80**	**284**	**1.76**	**295**	**1.91**	**240**	**0.26**	**295**
	4.66	*9615*	*3.21*	*3068*	*1.44*	*520*	*1.25*	*459*	*0.56*	*76*
Not on Treatment	**30.16**	**95**	**10.70**	**92**	**1.78**	**95**	**1.77**	**95**	***0.00***	***0***
	4.95	2865	3.73	984	1.40	169	1.37	168	*0.00*	*0*

Fig. 4: Birth Trends: Total, Intended and Unintended



In the era of HIV/AIDS and HAART, reproductive behaviors is a matter of public health which needs to be separated from risky sexual behavior. Indeed, about 75% of patients declared to desire a child within the next two years maximum, 20% responded not to desire any, and the remaining want a child in the near future. In the majority of the cases, among the reasons reported for not wanting a child was marital status or lack of a partner. Disclosure of HIV status and couples counseling was indeed another threat to public health. Independently from marital status, partner's HIV status is unknown to the patient in the majority of the cases, as well as her status to the partner. Subjective perception of future life expectancy among the patients is crucial in reproductive health. To a question formulated in the following manner: "*According to you, by how much is HAART known to prolong life expectancy of people living with HIV/AIDS*?". They had to choose between different options: a) 15-20 years or more, b) 10-15 years, c) 5-10 years? and d) I don't know. I find a statistical difference between women who were on treatment and those who did not yet start treatment. A higher percentage of women on treatment chose the option *"15-20" or more. A*mong those not yet on treatment they revealed higher uncertainty by choosing the option *"I don't know"*. Furthermore, to the question if AIDS was like any other chronic disease, I found women on treatment were more likely to choose the yes option. More details are provided in Table 13.

Tab. 13: Mean Comparison between Treated and Non-Treated on Perception of HIV/AIDS

	Mean					Observations	
Variable	Non Treated	Treated	Difference		P-Value	Non Treated	Treated
1 Good Knowledge of PMTCT	0.644	0.620	-0.024		0.336	101	292
2 Knows Someone With AIDS	0.530	0.735	0.204	***	1.000	100	290
3 Life expectancy 15-20	0.610	0.736	0.126	***	0.992	100	292
4 Certain about future Life Expectancy	0.700	0.788	0.088	**	0.963	100	292
5 Perceive AIDS as a chronic Disease	0.812	0.877	0.065	*	0.947	101	292
6 Certain about AIDS Chronic/not Chroni	0.891	0.949	0.058	**	0.977	101	292

*** p<0.01, ** p<0.05, * p<0.1.

Note: The Variables: "Certain about future Life Expectancy" and "Certain about AIDS Chronic/not Chronic" are complement to the answer "I don't know". Women on treatment are less likely to take the option "I don't know" on their future life expectancy and whether AIDS is a chronic disease. Robust standard errors are clustered at the enumeration area level.

3.2 Empirical Strategy

Treatment initiation is determined by an interaction of individual's CD-4 count and a random shock due to any kind of infectious disease. As discussed above, assignment to treatment results from either eligibility based on CD-4 counts (disease progression to AIDS) or critical health status (co-infection with other diseases like for example tuberculosis or hepatitis). CD4 counts are unlikely to be directly self-monitored by patients but can be a result of unobservable or observable individual characteristics like living standards.

PMTCT programs induce selection on pregnancy as women discover their status, level of CD4 count and treatment availability with pregnancy. Reverse causality puts in cause the internal validity of inference as pregnancy among HIV positive women induces treatment and vice-versus. With the PMTCT program, if the woman's disease has not progressed to AIDS, she is eligible for treatment *only* three months before and after child delivery. The analysis of this chapter is based on in treatment initiation due to progression to AIDS. Moreover, the Day Hospital of Yaounde is not specific to PMTCT so women on treatment are enrolled for several reasons and are likely to differ on many aspects. In any case, in the main analysis, I excluding women who started treatment or discovered their HIV status in the same year they had a child. The total amount of patients excluded are 92. Finally, PMTCT bias downwards the parameters of interest and does not affect the conclusion of the chapter as those women are less likely to bear a child after one year of treatment. Our results are based on the probability of childbearing in *"t+1"* given the women started treatment in *"t"*. Those

women had a child in *"t"* so this makes them unlikely to bear a child in *"t+1"*.

The analysis is based on time variation of treatment and pregnancy. I measure if change in fertility is associated with treatment initiation by controlling for unobserved individual and time fixed effects. Initially, the empirical strategy is based on the so called *moving window* where I consider two treatment periods.[24] I compare the average probability of giving birth in period *'t+1'* compared to period *'t'* for an individual who begins treatment in period *'t'*. Controls for these individuals are those who start treatment in period '*t+2*' where I look at their behavior in "*t*" and "*t+1*".[25] Each group of individuals who start treatment in different years have their own controls within a data-set constructed with two time periods: "*pre* " and "*post*". The *moving window* implies replacement of these controls and treated in the database. Some individuals are considered controls or treated at different points in time. In the study, patients are not replaced more than twice: as controls and treated. I restrict controls to individuals who start treatment in t+2 such that treated and controls are as comparable as possible. The choice of "*t*" and "*t+1*" is motivated by biological considerations: patients react to HAART and give birth after one year of treatment. If they conceive a child in the same year they started treatment, they are likely to give birth the year after. Individuals might also start treatment after conception and in this case the parameter of interest will be downward biased but the main conclusion of the paper will not be affected. Upon the constructed data-set, I use difference-in-differences approach, extended by adding a vector of individual characteristics to control for differences in observables between the groups at baseline. I estimate regression of the following form:

$$outcome_i = \beta_0 + \beta_1 ART_i + \beta_2 Post + \beta_3 ART_i * Post + \pi_k \mathbf{X}_i + \tau_t + \varepsilon_i \qquad (7)$$

where outcome is birth of individual i in period t (or intended-birth); ART is a treatment dummy if individual 'i' started treatment in year t; Post is the post-treatment period, i.e. a *'t+1'* dummy constructed from the moving window. Finally, τ_t represent year dummies and $\mathbf{X}_i$ is a set of observable characteristics which includes age, education and childbearing in *'t-1'*. The empirical strategy based on moving window

[24] This method has been applied in labor economics by Monteiro (2004) in estimating the effect of privatization on wages in Portugal. It has first been used in Kluve *et al.* (1999). A good advantage of the moving window is the short period of analysis pre-post, which avoids auto-correlation in the error terms and allows to control for baseline characteristics that are time invariant.

[25] I also replicate the results by looking at behavior in*'t-1'* as compared to *'t+1'*. The reason why I focus on *'t'* rather than on *'t-1'* is to avoid loss of observations because individuals who start treatment in 2003 cannot be evaluated. Moreover, patients who start treatment in *'t+2' are physically better off in 't-2'* and cannot serve as controls for those treated in *'t'*.

allows to control for individual fixed effects, to account for unobserved heterogeneity constant over time, and for time-invariant variation across individuals. However, the moving-window analysis does not control for different trends in the treatment and control groups.[26] To this purpose , I further extend the analysis with panel data-set on a larger window, from 2003-2010 and identify the treatment effect from variation within the individuals over time. Additionally, I control for individual specific trend that vary across individuals. It represents a flexible way to control for heterogeneous pregnancy behavior. I also allow the trend to depend on treatment initiation: in addition to shifting the level of outcome, treatment also affects the trend.

$$outcome_{it} = \theta_i + \tau_t + \theta_i t + \beta_1 ART_{it-1} + \beta_2 ART_{it-1} * t + \beta_{3k} \mathbf{X}_{it} + \varepsilon_{it} \tag{8}$$

3.3 Impact of Treatment on Pregnancy

Results from the Moving-Window

Table 14 shows results from the estimation of treatment on total and intended pregnancy based on Equation (7). One year of treatment is likely to increase the probability of childbearing on average by 6.36 % points compared to women who start treatment 2 years later. Very similar patterns with slightly higher coefficients are observed in column 2 and 4 for intentionally conceived children *i.e.* 7.3 % points. It is important to control for age, education and birth in period *"t-1"* due to upward bias in the parameter of interest.

[26] It is based on the Diff-in-Diff approach, which assumes parallel trend between the two groups.

Tab. 14: Coefficients from DIFFinDIFF based on the Moving Window: Birth in t+1 compared to t

Dependent Variable	Total Pregnancy	Desired	Total Pregnancy	Desired
	(1)	(2)	(3)	(4)
Post	0.0132	-0.0132	0.0219	-0.00834
	(0.0224)	(0.0197)	(0.0247)	(0.0217)
ART	-0.0759***	-0.0693***	-0.0676***	-0.0636***
	(0.0152)	(0.0146)	(0.0187)	(0.0174)
Post*ART	0.0852***	0.0870***	0.0636*	0.0730**
	(0.0294)	(0.0259)	(0.0344)	(0.0304)
Age			0.0244	0.0290**
			(0.0152)	(0.0118)
Age sq			-0.000508**	-0.000529***
			(0.000256)	(0.000203)
Lagged Child			-0.0484***	-0.0405***
			(0.0169)	(0.0134)
Education			-0.00616***	-0.00381**
			(0.00232)	(0.00189)
Year FE			YES	YES
Constant	0.0759***	0.0693***	-0.120	-0.264
	(0.0152)	(0.0146)	(0.217)	(0.162)
Observations	1,095	1,095	1,024	1,024
R-squared	0.022	0.016	0.045	0.030

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: The dependent variables in the first and third columns are total number of children born while in column two and four we refer to intended births.

I further decompose the average treatment effect on childbearing over the last seven years based on year of treatment initiation. I show the heterogeneity of the parameter in Table 15. Each column is a difference-in-difference estimator of treatment and control groups on desired pregnancy or else it is a sub-sample of the moving-window illustrated in Table 14. The magnitude and sign of one year treatment effect varies across patients. Interestingly, women who started treatment in earlier years were less likely to bear a child after one year of treatment while after the policy, women increasingly have intended pregnancy.[27]

[27] Lack of observations might render the results imprecise for inference. However, same results are found from panel data analysis.

Tab. 15: Dynamic of the Coefficients by Year of Treatment Initiation

	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010
Desired Pregnancy	**-0.147**	**-0.0333**	**0.144***	**0.129***	**0.0979***	**0.206*****
	(0.0954)	(0.0593)	(0.0768)	(0.0677)	(0.0544)	(0.0514)
Controls	YES	YES	YES	YES	YES	YES
Obs	86	100	152	244	226	444
Rsq	0.121	0.036	0.072	0.049	0.026	0.069

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Each column is a restricted sample of women who start treatment in different years with the restricted pre-post periods i.e. it is a decomposition of the moving window. Each regression includes socio-economic controls as in the moving window.

Results from Individual Specific Slopes depending on Treatment Initiation

In this paragraph I exploit information on a larger window as compared to the difference-in-difference conducted with the *moving-window* setup. In Table 16, I run different regressions on overall the sample, including women who were not yet on treatment in 2010. Column 1 shows results from a simple regression for ease of comparison while in Column 2, I control for fixed effects to capture unobserved heterogeneity at the individual level. The coefficient on treatment is quite sensitive to the propensity of childbearing of individuals that is constant over time. In Column 3 , year fixed effects reveal comparable results as in the moving window. Individual and year fixed effects explain a good part of the variability in pregnancy.

Column 4 allows for individual specific trends that capture unobserved factors influencing pregnancy at individual level to have a linear trend and allow this trend to vary across individuals. The coefficient have almost doubled: at least one year of treatment increases an individual's propensity to be pregnant by 15 percentage points. The specification allows individuals to have their own trend which can be increasing or decreasing depending on health status and age.

A 15 percentage point increase in pregnancy due to at least one year of treatment is relatively a very big impact and not only in absolute terms. Compared to the mean pregnancy rate of 0.14 in the sample, a coefficient of 15 percentage points increase corresponds to approximately a 107% increase in pregnancy rates, i.e. treatment has more than doubled the chance of pregnancy among HIV positive women, a very similar result found in epidemiology.

There is a possibility that the unobservables exhibit a more complex dynamic behavior than just a linear trend. I relax this assumption by separating out pre-existing trends from the dynamic response of a policy shock. In first place, I allow the trend to vary pre-post 2007 policy to investigate how treatment effect varies due to free access. Column 5 shows that women on treatment after 2007 are more likely to increase the average propensity of childbearing: 19.5 percentage points. This result is the average positive effect of the Cameroonian 2007 policy on free access to treatment on all treated women.

In Column 6, I decompose the policy effect year by year, as in Table 15. The coefficients are slightly higher and the patterns are very similar. As in the above regressions, the coefficient of treatment effect on pregnancy increases over time compared to the baseline. This is intuitive as HIV infection is increasingly perceived as a chronic dis-

Tab. 16: Coefficients Estimates of Treatment on Pregnancy

LPM	Dependent Variable: Pregnancy					
	(1)	(2)	(3)	(4)	(5)	(6)
ART	0.0166	0.0359**	*0.0826****	*0.159****	-0.0111	-0.0404
	(0.0141)	(0.0161)	(0.0195)	*(0.0274)*	(0.0159)	(0.0259)
POLICY					0.00353	
					(0.0340)	
POLICY*ART					*0.195****	
					(0.0368)	
ART*2005						0.0756
						(0.0542)
ART*2006						0.0110
						(0.0406)
ART*2007						*0.224****
						(0.0859)
ART*2008						*0.265****
						(0.0923)
ART*2009						*0.174***
						(0.0814)
ART*2010						*0.255****
						(0.0535)
IND FE		YES	YES	YES	YES	YES
YEAR FE			YES	YES		YES
IND SPEC TR				YES	YES	YES
LINEAR TREND					YES	
Constant	0.102***	0.0985***	0.106***	-0.0689***	0.0332	0.0179
	(0.00636)	(0.00304)	(0.0143)	(0.0224)	(0.0236)	(0.0222)
Observations	3,174	3,174	3,174	2,780	2,780	2,780
R-squared	0.000	0.002	0.008	0.010	0.010	0.012
Number of id		407	407	405	405	405

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: In Column 1 we report results from a simple regression model while in Column 2, Column 3 and Column 4, we respectively control for individual fixed effects, year fixed effects and individual specific trends. In Column 5 and Column 6, we allow the trend to vary by treatment status. In Column 6, we evaluate the 2007 policy while in the last column we evaluate the policy year by year.

ease allowing individuals to conduct normal lives.

To sum up, there is a positive effect of treatment on pregnancy which is heterogeneous by year of treatment. Women enrolled after 2007 are likely to bear a child as compared to those before the policy. Different mechanisms might explain this finding and they are discussed in Section3.6.

3.4 Heterogeneous Treatment Effects

Treatment Effect by the Number of Children Before treatment Initiation

Information on pre-treatment number of children enables to understand if treatment and the 2007 policy affected total fertility rate of women living in Yaounde. I estimate the coefficients by interacting treatment with pre-treatment number of children. The total number of children pre-treatment that is invariant over time is captured by individual fixed effects. Results are shown in Table 17.

Tab. 17: Coefficient Estimates by Number of Pre-Treatment Children

Heterogeneity by Pre-Treatment Number of Children		
Dependent Variable	Pregnancy	
	(1)	(2)
ART	0.0826***	0.120***
	(0.0195)	(0.0272)
ART*One Child.		-0.0465
		(0.0350)
ART*Two Child.		-0.0510
		(0.0402)
ART*Three Child.		0.0197
		(0.0775)
ART*Four Child.		-0.0847
		(0.117)
ART*Five Child.		-0.276**
		(0.115)
ART*Six Child.		-0.280***
		(0.0411)
ART*Seven Child.		-0.340***
		(0.0285)
Constant	0.106***	0.106***
	(0.0143)	(0.0143)
Observations	3,174	3,161
R-squared	0.008	0.010
Number of id	407	405
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1		

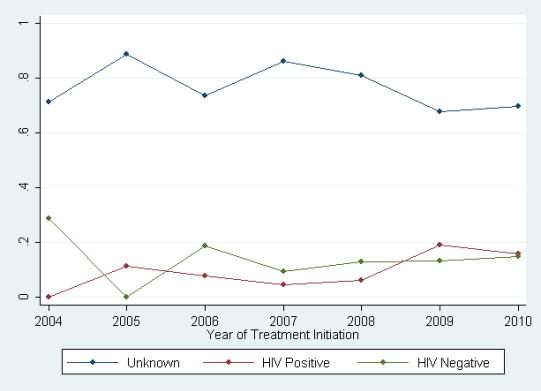
Note: Excluded categories are women who started treatment with seven or more children before treatment initiation. The analysis is based on year and individual fixed effects. Standard errors are clustered at the individual area level.

The coefficients in Table 17 suggest that treatment affects most pregnancy among women who start treatment without children as compared women not on treatment. Given women in Yaounde have three children on average (DHS 2004), HAART does not seem to affect those with positive number of children. It increases the propensity by twelve percentage points among those women with respect to women not on treatment. The effect decreases with additional number of children pre-treatment and assumes a negative coefficient for women who had four or above children before treatment initiation.

This finding is consistent with the idea treatment does not increase total fertility rate in the population as those without children or lesser than the average are most affected by HAART. In fact, women who had already reached the local average fertility rate are not affected by treatment. This suggests that treatment allows women to reach their desired number of children and does not lead to an increase in the latter. In line with this finding, the negative coefficient on ART in Table 14 withing the analysis of difference-in-difference suggests that HIV positive women who start treatment in a short delay have lower number of children as compared to those who start treatment after a longer period. The increase in birth is observed just after treatment initiation suggesting HIV positive women are more likely to delay birth after treatment initiation.

There exists a possibility of heterogeneous impact of treatment depending on marital status. Marriage patterns differ a lot from the time of registration at the hospital until the date of interview in 2010. Bargaining power among couples and intra-household bargaining might be explaining the results depending on marital status of women interviewed and awareness of their partner on their seropositivity. In Figure 5, I show how unawareness on partner's status prevails in married and cohabiting couples. Partners are unlikely to be aware of their partner's HIV status. Moreover, those women live in the capital city or near-by, with a good percentage of them having completed secondary school. This characteristic alleviates issues related to gender gap and intra-household fertility decisions. It is likely that intended pregnancies declared and observed by patients is reliable in measuring their own willingness in reproductive choices.

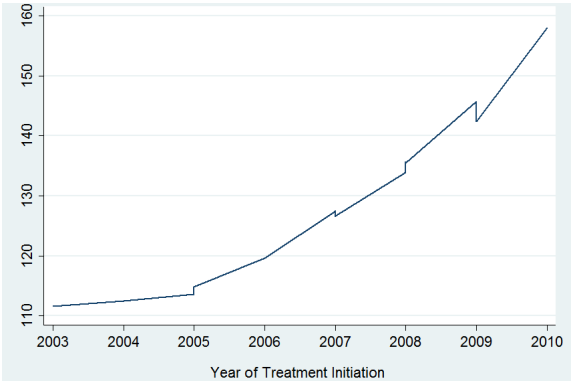
Fig. 5: Knowledge of Women on their Partner's Status



Treatment Effect by the level of CD-4 at Treatment Initiation

The literature in epidemiology states that there is no optimal time for treatment initiation. If there are no other co-infections the WHO recommended to start treatment with CD-4 below 200 cells/μl in 2010. From 2010, the threshold is raised to 350 cells/μl because "hit hard and early" strategy has been found to have increased survival. Recent studies also support earlier initiation of treatment - even before CD4 count drops below 350 cells/μl . Early anti-retroviral therapy has also a clinical benefit for both HIV positive individuals and their uninfected sexual partners suggesting the use of antiretroviral treatment as a part of a public health strategy to reduce the spread of the epidemics (Cohen *et al.*, 2011). In Cameroon, the average CD4 count at treatment initiation is increasing over time as shown in Figure 6 suggesting that people are increasing being screened and enrolled earlier when put on treatment.

Fig. 6: Lowess Estimate of CD4 at Treatment Initiation



Tab. 18: Random Individual Specific Trend dependent on Initial CD-4

Variables	Pregnancy		Intended Pregna	
	(1)	(2)	(3)	(
ART	0.159***	0.201***	0.109***	0.13
	(0.0274)	(0.0401)	(0.0223)	(0.0
(CD4<144)*ART		-0.0874*		-0.0
		(0.0500)		(0.0
YEAR FE	YES	YES	YES	YI
Constant	-0.0689***	-0.0699***	-0.0475**	-0.04
	(0.0224)	(0.0223)	(0.0184)	(0.0
Observations	2,780	2,780	2,780	2,7
R-squared	0.010	0.011	0.009	0.0
Number of id	405	405	405	4(

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Note: The variable CD4<144 takes the value 1 if CD4 is below the average. Results do not change if I consider the median of CD4 rather than the average. We test the hypothesis on total pregnancy and intended pregnancy. Robust standard errors are clustered at the individual level.

I test the hypothesis if response to pregnancy depends on CD-4 count at treatment initiation, *i.e.* if patients who start treatment earlier are more likely to be pregnant. Table 18 shows that women who start treatment earlier with CD-4 count above the average have higher propensity to pregnancy. In Column (1) I replicate results from the model where individual specific trends is controlled for i.e. Column (4) of Table 16. In Column (2) I interact treatment and a dummy of CD4 at treatment initiation is below the sample average. The coefficient on the interaction term between CD4 below the average and treatment initiation is negative and significant at ten percent. It is likely that women who start treatment with low values of CD4 count are worse off and less likely to conceive a child. Similar analysis is conducted in Column (3) and Column (4) where the dependent variable is intended pregnancy. The negative coefficient on the interaction term is not significant with pvalue that is slightly higher than 10%. Indeed a drawback with our analysis is variability in the dependent variable that drastically reduces power to detect the effect.

3.5 Threats to Internal Validity

Attrition

The above analysis is based on a random sample of individuals interviewed in the Day Hospital of Yaounde where information was collected retrospectively on their behavior in the past seven years. This implies that individuals enrolled in the program recently are oversampled. Indeed, I do not take into account women who might have died of AIDS at any point in time between 2003-2010 and were treated in the Day Hospital. The death of these patients at any given time in the window can be due to socio-economic characteristics affecting both fertility decisions, screening and monitoring of CD4 counts . To overcome this bias, I verify if results on a more recent window, between 2008-2010, give similar patterns as they are good representatives of the population on treatment. Every woman between the age of 15-40 with medical record registered from 2008 onwards was tracked by phone and interviewed at the hospital. Attrition due to death or drop-out is smaller in this window of 2008-2010. Column 1 of Table 19 shows the above results hold and the effect is bigger on average in terms of magnitude. This suggests that women who are on treatment for longer period are different and non comparable to those enrolled in recent year. They are likely to be better-off economically with lower propensity for pregnancy. I use three time periods to control for individual specific time specific effects and the model takes into account individual specific trends.

Endogeneity

In this section, I check if results are not driven by omitted correlates of treatment and pregnancy. There are time variant shocks that affect contemporaneously treatment initiation and pregnancy that are not captured by individual fixed effects or individual specific trends. I exploit the exogeneity of the Cameroonian policy on free scaling-up of treatment. To this aim, I restrict the sample to individuals who started treatment on and before 2007 such that individuals are comparable in health and socio-economic outcomes. There are only 98 individuals who were on treatment before the policy. Controlling for secular trend and individual fixed effects, the coefficient on pre-post 2007 should not have a direct effect on pregnancy. If the coefficient is significant it implies that women on treatment before 2007 were all affected by a common factor after the policy. In Column 2 of Table 19, I show that the effect of treatment depends on the 2007 policy although not statistically robust. This suggests that the increase in

Tab. 19: Sensitivity analysis: Attrition, Selection and Omitted Variables

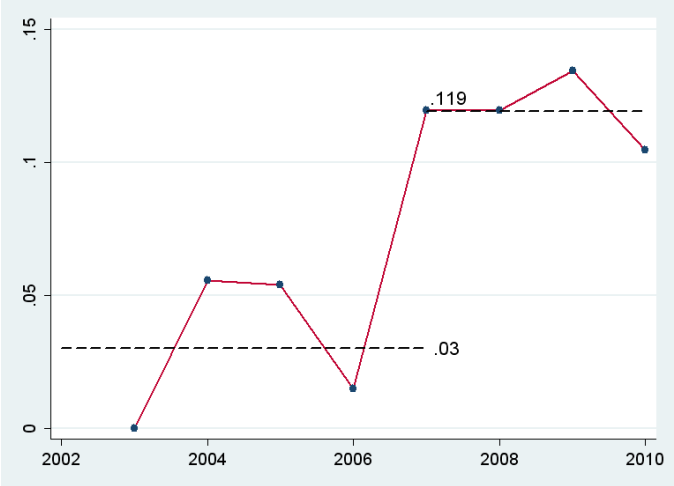
Dependent Variable	Pregnancy		
	(1)	(2)	(3)
ART	0.226***	0.00834	0.171***
	(0.0447)	(0.0278)	(0.0362)
ART*2007 Policy		0.178*	
		(0.0919)	
2007 Policy		0.0239	
		(0.0738)	
Trend		-0.00702	
		(0.0168)	
YEAR FE	YES		YES
CLASSE AGE	YES		YES
Constant	-0.0824	0.00296	-0.0523
	(0.103)	(0.0492)	(0.0799)
Observations	1,186	684	2,034
R-squared	0.032	0.017	0.012
Number of id	400	98	297

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: In Column 1, I show the effect of treatment on the restricted sample between 2008-2010 in order to avoid bias due to attrition. In Column 2, I show results on the restricted sample of individuals who start treatment before 2007 and evaluate the impact of the policy as an exogenous variation. In Column 3, we restrict the sample to individuals who start treatment in the same year they discovered their status. Standard errors are clustered at the individual area level.

pregnancy is partly driven by free access to ART, at least for these individuals, which is not determined by other unobservables related to health status. As patients were already on treatment before 2007, the coefficient captures the change in their behavior due to free scaling-up of HAART. Figure 7 shows the average number of births per year of 67 individuals who were already on treatment in 2006. The average number of children born from these women is increasing over time with the big jump being characterized after the 2007 policy. The difference is statistically significant and these individuals are comparable in terms of health, years of infection, treatment and socio-economic characteristics, which were all subject to similar social and economic costs of infection.

Fig. 7: Individuals on Treatment Before 2006



Reverse causality could also confound the estimates in Table 18. Self selection of healthy individuals with CD4 above the average might result from reproductive choices. As such, the relationship between treatment and pregnancy could be a result of planned pregnancy and treatment initiation. I test if results hold on a restricted sample of individuals who started treatment at the same time as were detected HIV positive. The amount of people who started treatment the same year of HIV detection amounts to a little less than 300 women suggesting Voluntary Counseling and Testing is not widespread. Results in Column (3) of Table 19 shows that previous patterns are not altered by self selection. This also rules out the case if individuals anticipated the 2007 policy and hence adapted their reproductive choices before treatment initiation.[28]

[28] In any case, the main conclusions of the paper will not be altered because anticipation will downward

3.6 Discussion and Concluding Remarks

This chapter uses first hand data collected among patients in the Central Hospital of Yaounde-Cameroon. It examines fertility response to a change in HAART availability, a treatment which provides enormous mortality benefits to HIV positive women. I use information on birth history collected retrospectively, date of treatment initiation, HIV test and other socio-economic variables. The staggered timing of treatment initiation allows to study fertility response between 2003-2010 by taking into account different types of heterogeneity: time variant and time invariant.

Over the window of analysis, treatment is associated with an increase in pregnancy by 0.15 on average after one year of treatment, which is a big effect both in absolute and relative terms. A 15 percentage points increase corresponds to doubling pregnancy rate among women living with HIV/AIDS. This result is very similar to other findings in epidemiology. I also find that treatment effect on pregnancy is time dependent. Compared to individuals who started treatment in 2004, HAART increases pregnancy response by 22.4 percentage points among women who started treatment after the 2007 policy. The coefficient increases over year of treatment initiation suggesting behavioral response that goes beyond the direct effect of health benefits. Also, treatment effect is heterogeneous by pre-treatment number of children: pregnancy response is highest among women who had zero number of children before treatment initiation while non significant or negative response is found among women who started treatment with at least one child. Given the average fertility per women living in Yaounde was 3.2 in 2004 (DHS survey), it is less likely treatment increases total fertility rate in the population. Women who started treatment with children above 3 have a negative coefficient as compared to those without any child. HAART restores fertility in women to (partially) reach the desired fertility rate. Furthermore, I find that treatment increased pregnancy among women who had higher initial CD-4 count which is partly explained by the fact that women start treatment earlier in recent years (Figure 6). To sum up, results suggest behavioral response beyond the immediate impact on health for the following reasons: *first*, the coefficient on treatment is increasing over time with a jump being characterized in 2007; *second,* treatment affects significantly a sub-population of individuals with zero number of children before treatment initiation*; third*, results are consistent when women who got tested or started treatment for reasons related to amenorrhea or menstrual disorder are excluded (there are few of

bias the treatment effect.

them as illustrated in Table 11); *fourth*, individuals who started treatment before 2007 increased their pregnancy after the policy.

Possible different mechanisms can explain the above results. The 2007 policy might have incentivized Voluntary Counseling and Testing (VCT) allowing individuals to better monitor their CD-4 due to the possibility of free access to HAART. Women starting treatment with higher CD-4 had better health outcomes and were more fecund after a shorter period. Early screening can further be explained by the decreasing social and psychological costs from infection. The cost associated to stigma, self-exclusion and trauma of infection are decreasing over time and increasing number of people are on treatment allowing patients to live a better life with the infection. Free access to treatment also contributes in early screening of individuals and their reproductive behavior due to income effect. Indeed, the 2007 policy signals indirectly long term availability and affordability. From descriptive statistics, women on treatment are likely to have a positive perception on their future life expectancy as shown in Table 13.

It might also be the case that counseling and sexual education provided by health workers has changed over time. In order to avoid attrition and control the epidemics, health worker are increasingly educating people living with HIV that HIV/AIDS is a chronic diseases as opposed to a mortal disease. This is likely to reduce psychological cost of infection and encourage people to follow-up with their therapy and reduce problems relative to drug resistance and viral mutation. In the past, HIV positive women who became pregnant were likely to be object of criticism and stigma by health workers.

Other possible mechanism is lower viral load and altruistic behavior of individuals who delay birth in order not to infect own partner. However, only 35% of women were aware of the fact that HAART reduces viral load and the treatment coefficient is not heterogeneous by this knowledge suggesting that altruism is less likely to drive the findings. Difference in socio-economic characteristics between those who started treatment in earlier years vs those in recent years is also less likely to explain the above results. The treatment coefficient on the 2007 policy as a quasi-experiment on those who started treatment before 2006 illustrates that reproductive choices were affected by income effect i.e. free access to treatment (Table 19).

The epidemiological implication of treatment availability and pregnancy in terms of HIV incidence is not the object of this study. However, an important implication of the analysis is the behavioral response of treatment on intended (or unintended)

pregnancy, which is reflected on unprotected sex . If treatment adherence is well developed among patients, the probability of HIV infection is dramatically reduced: both horizontally (partner) and vertically (child). The fact that pregnancy response is higher among women who start treatment with CD-4 higher than the average might suggest that pregnancy occurs among women with lower viral load and infectivity. In terms of policy implication, a more comprehensive approach that includes family-planning and couples counseling is fundamental with treatment availability.

4 Sex Composition of Older Siblings and Teen-Pregnancy in Patriarchal Communities. Evidence from Central Africa

Patriarchal societies are determined by a well defined power and kin relationships based on gender. Men often have control of the public sphere that legitimate their role and define women's status, privileges and rights in the society. Many communities in Sub-Saharan Africa are characterized by this form of social organization, although with different levels of gender related power and role. Indeed, such gendered power system is often based on a network of social, political and economic relationships where men dominate and control female labor, reproduction and sexuality. This system has been intergenerationally transferred through the capacity of fathers to bequeath to their sons the power to control and monitor female members of the household when young and command the resources and operate over the labor force of their wives and children when old. In such a set-up, it is likely sex composition of siblings has a direct affect on women's well-being within the household by factors that are not driven by parental choices or allocations. This chapter aims to address this particular characteristics that sibling's sex composition and gender roles have on teen pregnancy in Sub-Saharan Africa (SSA).

Teen pregnancy is still prevalent in Sub Saharan Africa. Adolescent fertility rate is around 123 live births per 1000 women aged 15-19 years (United Nation, 2009) and teenage pregnancy rate is over 20% (ICF,2012) in the region. Teenage pregnancy and early childbearing generate important socioeconomic consequences and can shape and alter the entire future life of the teen (United Nation, 2009; Panday *et al.,* 2009; Singh 1998; Geronimus and Korenman 1992). The determinants of risky sexual behavior of teens, and hence teen pregnancy, have not been fully elaborated. The literature on public health suggests a set of factors that affect (risky) sexual behavior. These factors are organized around different levels: individual characteristics, group characteristics, institutional and legal framework. Panday *et al.* (2009) provide a review of these factors and discuss how they are interrelated.

Recent papers in economics have shown that birth order of siblings matter for risky behavior of adolescents. Given background characteristics, later born siblings are more risk takers than firstborns in terms of consuming tobacco, alcohol, marijuana and engaging in risky sexual behavior (Ouyang 2004; Argys *et al.,* 2006; Averette *et al.,* 2011). The main channel through which these effects prevail is parental supervision: firstborns are closely supervised by their parents as compared to later born and the

increased level of supervision is associated with lower risk. Averette *et al.* (2011) conclude that peer effect prevails beyond the parental supervision as the effect of older siblings on younger ones persist even after controlling for parental supervision.

This chapter focuses on men's control over sexuality and reproductive behavior of women in the above described patriarchal context. Given the majority of SSA societies are patriarchal, this paper shows how older brothers exercise direct control over sister's sexuality and behavior within the household, as compared to other siblings. In developing societies, older siblings have care-taking roles towards the younger ones who are often supervised by them for extended periods while parents are away or at work. Younger siblings are often taught to respect their older siblings and cooperate; the relationship between them often end up interdependent and hierarchical, with responsibilities and authorities like that of parents. Brothers exercise direct control over their sibling's sexuality and have better comparative advantage as watchdogs towards bad pretenders because of age distance from younger female teens of the household and their potential sexual partners. They can easily screen partners and monitor the behavior of sisters in order to avoid out-of-wedlock pregnancy in the family. This paper attempts to evaluate this role of male firstborns on risky sexual behavior of teen sisters.

To my knowledge, this is the first research that looks at how gender inequality shapes sexuality and reproductive behavior of adult women with long-term consequence on their well-being. Indeed gender inequality in income, education, health, inheritance and labor supply have been extensively analyzed in the economics literature. It contributes to the existing literature in two different ways. Firstly, how siblings affect each other in future economic outcomes and well-being across many dimensions. In this respect, birth order and gender are found to be important determinants of sibling's well-being. Birth order has a role on risky behavior of younger siblings through two channels: parental supervision and peer effect. This chapter brings forth this literature by investigating the interaction of birth order and gender on younger sibling's behavior. Secondly, there is a growing literature on son preference in SSA, in line with the analysis of missing women in Asia.[29] Different motivations have been put forth to explain gender bias in fertility preference in SSA, among which patrilineality of property inheritance and especially that of land. [30] This chapter extends this lit-

[29] For missing women in Asia, I refer to the extensive literature on sex ratio and selective abortion brought forth by different authors like Sen (1990). For sex preference in SSA, I refer to Anderson and Ray (2010) and Milazzo (2012).

[30] There is an extensive literature on how marriage, inheritance and institutions affect women's well-

erature by providing further explanation of the *"demand for son"* among women in Africa. Male members of the family, and specially older ones, contribute in serving as watchdogs for younger female siblings for their transition to adulthood. Not only they reduce direct family-costs from being pregnant, they also create positive externalities for members of the household by reducing costs associated to premarital pregnancies of sisters. The chapter sheds light on the advantage of sons in strong patriarchal organizations by shaping sexuality and reproductive behavior of young women.

DHS data for Cameroon, Nigeria and Chad is used to investigate the effect of gender composition of older siblings on adolescent's risky sexual behavior in SSA. Indeed, in SSA context where contraception use is low and abortion is illegal and unsafely practiced, engaging in risky sexual behavior can be very costly for the teen and the family. Out-of-wedlock pregnancy and premarital sex among teen girls is undesired by parents who want to avoid the associated costs. The findings of this chapter suggest that teen girls are less likely to engage in deviating behaviors if born in male firstborn families where an older brother functions as a watchdog in their transition to adulthood. They are roughly 2.5 percentage points less likely to have had premarital childbearing with respect to peers in female firstborn families. In relative terms, this corresponds to almost 50 percent reduction. Survival analysis on age at first birth also suggests older brothers delay the timing of out-of-wedlock pregnancy giving evidence of teen's protection in their transition to adulthood. The analysis is widened to investigate the difference between male firstborns and any male older bother.

The identification strategy is based on a simple "natural experiment" within the household. Parents decide to have a child but do not choose gender of their firstborn. Thus gender of the firstborn is exogenous in a household decision making and behavioral model. Although, exogenous at birth, female teens from male firstborn families might still systematically differ from those in female firstborn families in older ages if sex of firstborn affects household characteristics. I estimate the parameters by controlling for these observables. Overall, comparison of secondborn girls in male firstborn families vs female firstborn families gives a causal interpretation in SSA context where almost all families have at least a second child.

Teen pregnancy has everlasting consequences. I also examine these consequences by analyzing the welfare status of women with out-of-wedlock pregnancy while teen in older ages. After controlling for several factors, having had premarital teen pregnancy among never married is associated with higher probability of being head of

being in SSA. See for example Guyer (1987).

the household rather than part of a larger family. Among those ever married, having had a premarital child while teen is associated with lower education of the husband and higher husband-wife age-gap. A child born out-of-wedlock from a teen mother is associated with fostering and being dead at the time of the survey.

The chapter starts by drawing a general picture on age based hierarchy and gender bias among siblings in patriarchal SSA. It discusses teen's social life and attitudes of parents towards premarital sex in Section 4.1. In Section 4.2, I describe data and empirical strategy while results are presented in Section 4.3. I address, in Section 4.4, mechanisms and alternative confounding factors on the role of male brothers as "*watch-dogs*". Finally, Section4.5examines potential welfare consequences of teen pregnancy and concluding remarks are laid in Section 4.6.

4.1 Age and Gender Based Roles among Teen Siblings in SSA Context

Texts in anthropology and sociology have shown how age-based hierarchy is widespread in SSA. Older siblings have responsibilities towards the younger ones which is accompanied by some form of authority. For example, Heritier (1981) discuss this factor: « .... *Le rapport aîné/cadet, même lorsqu'il s'exerce entre hommes, peut être traduit en termes de génération, comme s'il s'agissait d'un rapport de père fils et non un rapports entre frères* » (F Heritier 1981 cited by Abèlès and Collard 1985, p207). Translated: « *...The relationship between firstborn/younger siblings, even among men only, can be interpreted in terms of generation, as if the relation is between father/son and not among brothers* ». The author puts emphasis on age based hierarchy and parenting among siblings in many in patriarchal societies of SSA countries where other than age, sex also constitutes authority. In fact, in Abèlès *et al.* (1985): «.... *l'ordre sexuel prime : si deux jumeaux de sexes différents naissent et que la fille se présente la première, on pense que le garçon va mourir parce qu'il ne pourra pas supporter l'offense; même s'ils sont plus jeunes que leurs sœurs, les garçons sont toujours servis en premier.*» (Abèlès et Collard 1985, p208). Translated: *".... gender based order dominates: if twins of different sexes are born and the daughter comes out first, it is believed that the boy will die because he cannot bear the insult; even when they are younger than their sisters, boys are always served first* ". These forms of hierarchy can also be seen in terms of land inheritence (including livestock): « *En ce qui concerne les animaux, l'aîné reçois trois fois plus que les autres fils aînés des autres épouses du père...* » (Abèlès et Collard 1985, p210); "*Regarding animals, the*

eldest boy receive three times more than firstborns of the other wives of the father".

In domestic relations, older siblings endorse protective behavior with responsibilities towards younger siblings. They often have similar authorities like that of parents. In Cicirelli (1995), the author underlines that care-giving of siblings serve as a backup system in the event that parents do not survive up to a certain age. The relationship is more than just custodial because they are combined with an educative mission to socialize and train younger siblings to become functioning members of the society. Indeed, younger siblings are taught to respect and obey older siblings as they would have to a parent (Cicrelle, 1995). The relationship between brothers and sisters also have an important role in marriage negotiations as they depend on one another to generate the wealth necessary for building own families and increasing household's wealth. Gender combined with age creates a different form of authority; male firstborns are the most dominant figures with the greatest seniority among siblings. « *...dans le contexte des relations des individus considérés dans leur appartenance à des unités domestiques, celui qui domines est l'aîné, celui qui est dominé est le cadet.....Il n'y a d'aîné et de cadet que d'hommes ; les femmes apparaissent comme des instruments de domination des ainés sur les cadets.* » (Marc-Eric Gruénais 1985, pg 221)[31]. Translating: " in the context of intra-household relationship, the ruler is the eldest son, he dominates the younger siblings ..... Elder and younger sibling's relationship exist only among males; women are seen as instruments of domination by firstborn". This implies that gender of the firstborn have important influences in terms of authority within the household and among siblinings.

Premarital teen pregnancy is undesired by the teen and parents for several reasons and importantly the associated economic costs: *first*, opportunity cost from schooling if the likelihood of attending school after puberty is low ; *second*: the parenting time and economic cost of raising the child; *third*, being forced into early marriage with partner when possible; *fourth*: higher difficulties in finding other partners who will take her and the child in charge and hence end up being fostered or single parent; *fifth:* those teens who value the return from education, they might decide to undergo an abortion which is often illegal and unsafe in developing countries. Last but not least, adolescent pregnancy is found to be associated with higher rates of morbidity and mortality for the mother and the child. Other costs which cannot be valued are those associated with the well-being of the child who is raised without the presence of

[31] Marc-Eric Gruénais 1985, « Aînés, aînées ; cadets, cadettes les relations ainés/cadets chez les Mossi du centre» dans A&C1985)

both parents.

Contraceptive use in Central Africa is very low and, because the cost of teen pregnancy is very high, parents and older siblings tend to control premarital sex of their teen girls. *"Parents exercise control over teenagers. The widespread belief is that a teenager's bad behavior may damage the families name and reputation. Parents tend to be harder on girls, always warning them to avoid premarital sex and manners that would cost them prospective husbands"* (Teen life in Africa. Toyin Falola,2004-p4). "Social life is active as teenagers meet one another to play and share experiences. Not all cultures encourage dating. *Boys have more freedom than girls and tend to socialize more outside of their homes and with many more people. Parents are always anxious to know the friends with whom their daughter socializes. In cultures where virginity is much valued, socialization can lead to premarital sex which is condemned"* (Teen life in Africa. Toyin Falola,2004-p6). As underlined by Falola (2004), parents tend to exercise direct control over their teen girls and together with older siblings they serve as "watchdogs" against non desired partners.

Male firstborns have an advantage in protecting younger female siblings from potential bad partners in their transition to adulthood. The main comparative advantage stems from two characteristics of male firstborns: gender role and age based hierarchy. *Firstly*, gender roles allow older brothers to socialize more outside the household and given networks are gender biased in patriarchal societies, they are likely to be closer to the network of potential partners as compared to older sisters. *Secondly*, in African societies, inter-generational sexual relations has been emphasized in the HIV/AIDS literature due to transactional sex.[32] Birth order allows male firstborns to be closer to the network of potential partners of younger sisters, as compared to younger brothers. Male firstborns also have higher authority and status within the household. In fact, from the domestic violence data-set, the percentage of teens who have been beaten by a brother are more than twice as those who have ever been beaten by a sister (the overall percentage is less than 10%). In conclusion, male firstborns are best positioned in serving as watchdogs for younger siblings.

[32] Prevalence of sugar daddies and risk of HIV/AIDS infection has been put in evidence in several studies. See for example Dupas *et al.* (2009)

4.2 Data and Empirical Strategy

Descriptive Statistics

I use the Maternal Mortality section module from the DHS surveys in three different patriarchal and bordering countries: Cameroon 2004, Nigeria 2005 and Chad 2005. The DHS is a representative sample of women in childbearing age (15-49) with different marital status. Individuals are asked to list all siblings from their biological mother, including their age, sex, mortality and number of childbearing. The choice of these countries is based on territorial comparability and cultural similarities. Indeed, transition from child-to-adulthood of teen girls, as well as gender roles, is specific to contexts and place of interview. [33]

Age at first birth is mainly concentrated between the age of 15-19, both within and out-of-wedlock pregnancy. Figure 8 shows the distribution of age at first birth by type of pregnancy; as expected, marriage delays age at first birth. Nonetheless, teen pregnancy is predominant in these countries.

Fig. 8: Age at First Birth: Within/Out-of-Wedlock

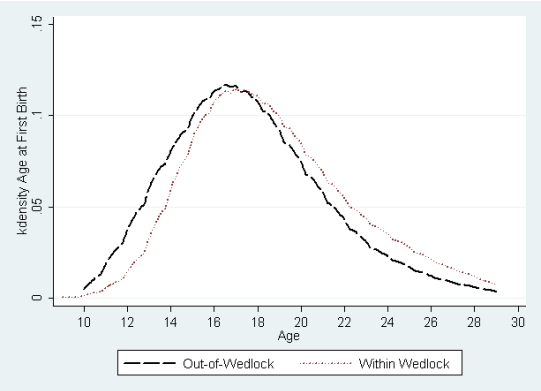


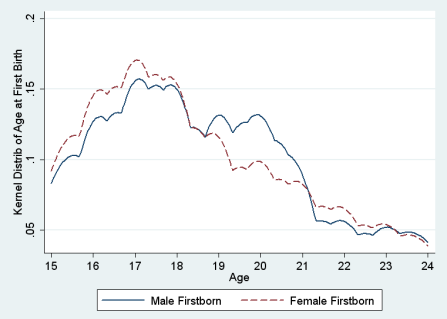
Table 20 shows descriptive statistics by country and generation. Although slightly different by country, women who had teen pregnancy are likely to give their first birth between the age of 16-17 across all cohorts in Panel A. Teens in Chad, endowed with lower level of schooling, bear a child earlier than their peers in Cameroon and Nigeria (Panel G). Similar pattern is observed for age at marriage; it is higher in countries with

[33] They reflect cultural values that drive marriage and fertility preferences, as well as economic development and access to schooling of girls. Education has played an important role in the timing of reproductive events among teens in Sub-Saharan Africa where access to primary and secondary school of girls has been increasing with different rates in the last 30 years.

higher level of girl's education. In Cameroon and Nigeria, the percentage of women who had a child or got married while teen is decreasing across cohorts while years of schooling is increasing (Panel B, E and G). However, one in two women bear a child before the age of 19 in Nigeria and it can reach up to 60% in Cameroon and Chad suggesting that women enter adulthood very early. A slightly different pattern is observed in Chad where schooling is improving across cohorts but remains relatively low. In the same line, polygamy is decreasing across cohorts, except in Chad, where a clear pattern is not found.

Premarital pregnancy is the lowest in Chad where transition from childhood to adulthood is relatively shorter decreasing the likelihood of premarital sex. Indeed, more than 80% of women get married before the age of 19. A good part of teen pregnancies occur within marriage but premarital pregnancy while teen remain relatively high as compared to other SSA countries. Cameroon has the highest rate of premarital pregnancy and with certain degree of stability across generations ranging between 12-16%. It decreases to 5-6% in Nigeria and 1-3% in Chad. Premarital childbearing while teen does not keep these girls out of the marriage market- more than 50% of them get married before the age of 19. The pattern is similar across cohorts and countries, though slightly higher in Chad. This might suggest that pregnant teens are forced into shot-gun marriage. I define shot-gun marriage as women who got married the same year they got pregnant or the year after. In Panel I, between 9-14% teens end up in shot-gun marriage with similar and comparable patterns across cohorts and countries. Estimation of shotgun marriage based on imprecise time-line is likely to be biased. Women who have been forced into marriage after the birth of the child are not taken into account.

Fig. 9: Kernel Distribution of Age at First Premarital Birth



Notes: I plot age at first birth of premarital pregnancies for women in male against female firstborns. The sample includes secondborn women aged between 15-24.

Tab. 20: Descriptive Statistics by Country and Cohorts

AGE GROUP	15-19	20-29	30-39	40-49	15-19	20-29	30-39	40-49	15-19	20-29	30-39	40-49
	PANEL A: TEEN AVERAGE AGE AT FIRST BIRTH				PANEL D: TEEN AVERAGE AT AGE FIRST MARR				PANEL G: YEARS OF EDUC			
Cameroon	16.41	16.65	16.61	16.68	15.30	15.88	15.64	15.59	6.01	6.21	5.43	4.32
sd	*1.76*	*1.81*	*1.81*	*1.77*	*1.77*	*2.12*	*2.16*	*2.04*	*3.26*	*4.09*	*4.08*	*3.83*
Obs	*772*	*2207*	*1474*	*935*	*880*	*2525*	*1644*	*1123*	*2679*	*3948*	*2439*	*1578*
Nigeria	16.22	16.51	16.33	16.21	*14.90*	15.44	15.22	15.08	6.30	6.37	5.27	3.66
sd	*1.59*	*1.88*	*1.95*	*1.96*	*1.71*	*2.23*	*2.19*	*2.22*	*4.43*	*5.52*	*5.48*	*5.05*
Obs	*1591*	*5740*	*4129*	*3122*	*2054*	*6902*	*5333*	*4092*	*6580*	*12397*	*8432*	*5944*
Tchad	16.19	16.46	16.25	16.24	14.90	15.19	15.07	15.02	2.59	2.11	1.15	0.86
sd	*1.59*	*1.86*	*1.93*	*1.92*	*1.71*	*2.15*	*2.09*	*2.08*	*3.53*	*3.83*	*2.74*	*2.50*
Obs	*474*	*1463*	*962*	*636*	*597*	*1807*	*1233*	*888*	*1459*	*2161*	*1415*	*1043*
	PANEL B: TEEN PREGNANCY (%)				PANEL E: TEEN MARRIAGE (%)				PANEL H: POLYGYNY(%)			
AGE GROUP	15-19	20-29	30-39	40-49	15-19	20-29	30-39	40-49	15-19	20-29	30-39	40-49
Cameroon	0.288	0.558	0.603	0.592	0.33	0.64	0.67	0.71	0.06	0.18	0.29	0.34
sd	*0.45*	*0.50*	*0.49*	*0.49*	*0.47*	*0.48*	*0.47*	*0.45*	*0.24*	*0.38*	*0.45*	*0.47*
Obs	*2680*	*3953*	*2444*	*1579*	*2680*	*3953*	*2444*	*1579*	*2680*	*3953*	*2444*	*1579*
Nigeria	0.241	0.463	0.489	0.525	0.31	0.56	0.63	0.69	0.08	0.20	0.34	0.39
sd	*0.43*	*0.50*	*0.50*	*0.50*	*0.46*	*0.50*	*0.48*	*0.46*	*0.27*	*0.40*	*0.47*	*0.49*
Obs	*6591*	*12406*	*8440*	*5948*	*6591*	*12406*	*8440*	*5948*	*6591*	*12406*	*8440*	*5948*
Tchad	0.325	0.676	0.678	0.610	0.41	0.84	0.87	0.85	0.08	0.29	0.41	0.36
sd	*0.47*	*0.47*	*0.47*	*0.49*	*0.49*	*0.37*	*0.34*	*0.36*	*0.27*	*0.45*	*0.49*	*0.48*
Obs	*1459*	*2164*	*1419*	*1043*	*1459*	*2164*	*1419*	*1043*	*1459*	*2164*	*1419*	*1043*
	PANEL C: TEEN PREMARRITAL PREGNANCY (%)				PANEL F: MARIAGE <20 AFTER TEEN-PREM-PREG (%)				PANEL I: TEEN PREG WITH SHOTGUN MARRIAGE (%)			
AGE GROUP	15-19	20-29	30-39	40-49	15-19	20-29	30-39	40-49	15-19	20-29	30-39	40-49
Cameroon	0.078	0.143	0.164	0.124	0.20	0.49	0.46	0.43	0.10	0.13	0.14	0.14
sd	*0.27*	*0.35*	*0.37*	*0.33*	*0.40*	*0.50*	*0.50*	*0.50*	*0.31*	*0.34*	*0.35*	*0.34*
Obs	*2680*	*3953*	*2444*	*1579*	*210*	*565*	*401*	*196*	*880*	*2525*	*1644*	*1123*
Nigeria	0.033	0.050	0.058	0.065	0.20	0.57	0.61	0.57	0.11	0.13	0.12	0.12
sd	*0.18*	*0.22*	*0.23*	*0.25*	*0.40*	*0.50*	*0.49*	*0.50*	*0.31*	*0.34*	*0.33*	*0.33*
Obs	*6591*	*12406*	*8440*	*5948*	*217*	*619*	*491*	*386*	*2054*	*6902*	*5333*	*4092*
Tchad	0.011	0.014	0.033	0.033	0.31	0.77	0.68	0.56	0.11	0.09	0.10	0.09
sd	*0.10*	*0.12*	*0.18*	*0.18*	*0.48*	*0.43*	*0.47*	*0.50*	*0.32*	*0.28*	*0.30*	*0.28*
Obs	*1459*	*2164*	*1419*	*1043*	*16*	*31*	*47*	*34*	*597*	*1807*	*1233*	*888*

Figure 9 shows how premarital pregnancy varies by gender of the firstborn. Kernel distribution of age at first premarital birth shows that women from male firstborn families are likely to delay out-of-wedlock pregnancy while teen. For empirical analysis, the exogeneity of firstborn's gender can be interpreted as a causal effect on second-borns and other confounding factors are unlikely to bias the parameter.

Women of different ranks in male vs female firstborn families are likely to bear systematic difference over time due to other household decisions: siblings rivalry, fertility stopping rule, marriage and labor market etc. Indeed, women from male firstborn families seem to be oversampled in Figure 10a. This might suggest that the two groups are not totally random and are likely to be subject to different environments. From Figure 10a, older women are likely to declare to be born in male firstborn families. The bias can be attributed to selection on gender of the firstborn affected by parental decision in fertility behavior models like "*the demand for son*".[34] Precisely, women from female firstborns end up with systematic higher number of siblings; they are likely to be worse-off due to resource scarcity and have lower probability of reaching older ages. Another explanation for the high percentage of male firstborn families stems from the fact that women with poor economic conditions have higher biological probability to give female births.[35] This gives further evidence that women from female firstborn families are generally from poorer socio-economic conditions. This implies a downward bias in the parameter of interest if these women are less likely to be sampled in a household based survey like that of the DHS or reach older ages.

Retrospective data suffer from data quality and bias related to miss-recalling and/or miss-reporting. The bias is likely to be correlated with socio-economic characteristics of individuals. [36] Recalling of birth history (own or that of the mother) suffer from gender based bias if women tend to recall an older dead brother (male birth) rather than an older sister (female birth) and specially if death of the sister (child) occurred in early ages.[37] Indeed, women from environments where the *"demand for son"* is

[34] Milazzo (2012) underlines that missing women exist in Nigeria due to social pressure for bearing sons. Son preference behavior in fertility puts women at risk due to maternal mortality. Indeed, the author finds that women with female firstbirth are more likely to have higher number of births, desire for addition children and shorter birth interval.

[35] Almond and Mazumder (2011) find a negative correlation between maternal malnutrition and male childbearing which is inherent to the weakness of male fetuses. The authors use Ramadan and date of birth to estimate this relationship.

[36] For a review on the type of bias related to retrospective data, Beckett *et al.* (2001) provide evidence from developing countries.

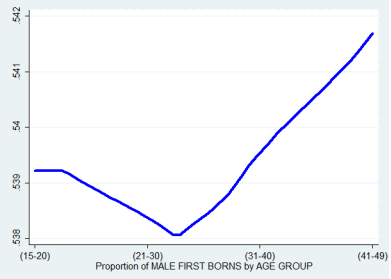
[37] Dyson (2001) provide evidence on miss-reporting of female births from India. Similar findings are found in China as well.

dominant tend to be gender biased when reporting birth history. In Figure 10b I plot the percentage of women from male firstborn families by total number of siblings and her rank. Given the rank of the woman, the proportion of male firstborns decreases with the increasing number of siblings. It confirms that male firstborn families are smaller on average. The figure also suggests that given the total number of siblings, the further the woman is from the firstborn, the more likely she declares to be from male firstborn family. In addition, individuals are likely to report a bias if death of the firstborn occurred in early ages. In Figure 10c I plot the proportion of dead male firstborns vs dead female firstborns by age at death. The proportion of male firstborns among those who died before reaching the age of one reaches 65% and gradually decreases when firstborns died after the age of 10. Recalling of older siblings seems to be gender biased where gender is negatively correlated with age at death. Women from poorer socio-economic characteristics are likely to be biased towards male firstborn families. This would further create a downward bias on the parameter of interest because women from female firstborn families are likely to be better off .[38]

[38] Recalling bias can also be due to biological weakness of male births who are likely to die before reaching the age of 1. However, I find that mortality of male births is particular to firstborns. Mean difference of survival status among siblings, given the firstborns are alive, suggests that women from male firstborns have lower mortality rate among siblings.

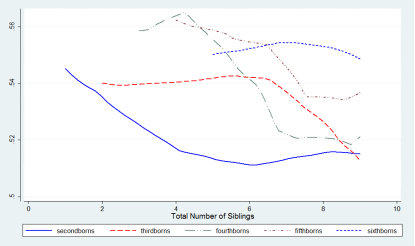
Fig. 10: Distribution of Male Firstborns

(a) **Male Firstborns by Age Group**



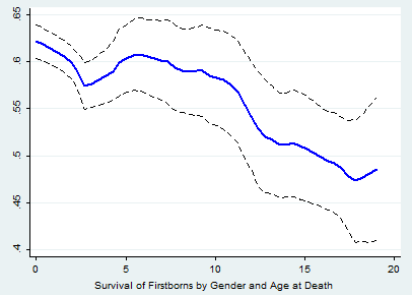
Notes: I plot percentage of women from Male firstborn families by Age Group. The sample includes all non first-born women aged between 15-49 with at least one sibling.

(b) **Male Firstborns by Preceding and Total Num of Siblings**



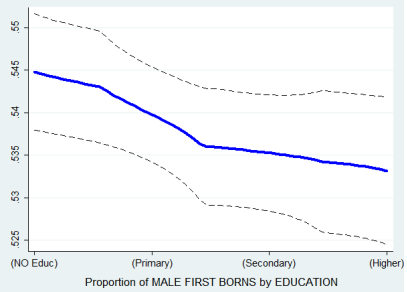
Notes: I plot percentage of women from Male firstborn families by the number of preceding siblings over total number of siblings. The sample includes all non firstborn women aged between 15-49 with at least one sibling -.

(c) **Dead Firstborns by Sex and Age at Death**



Notes: I use the sample of women with dead firstborns and plot the percentage of male firstborns against age at death. Again, the sample includes all non firstborn women-with at least one sibling -aged between 15-49.

(d) **Male Firstborns by Education Level**



Notes: I plot percentage of women from Male firstborn families against their Education Level. Again, the sample includes all non firstborn with at least one sibling -aged between 15-49.

All the above figures and discussions suggest that socio-economic conditions are likely to be correlated with gender of the firstborn among siblings. Indeed, in Figure 10d education of the women si correlated with gender of the firstborn. Women with lower level of education are more likely to report being from a family with male firstborn. Overall, findings suggest selection; socio-economic characteristics of the women need to be taken into account for the empirical analysis.

In Table 21, I test mean difference on socio-economic characteristics between the two groups. I further restrict the sample to women with firstborns alive to reduce bias due to recalling or biological survival of male births. I find that women from male firstborn families are generally older and worse-off in terms of wealth and years of education, although not statistically significant if survival of firstborn is not taken into account. The number of current household members and number of siblings also seem smaller and statistically significant if firstborn is alive. Overall, women from male firstborn families live in smaller families and survival of male firstborns is lower confirming high death rate among male births; interestingly, among firstborns alive, the number of siblings who died before reaching five is lower if from male firstborn family and it is statistically significant. This suggests mortality of firstborns before reaching the age of one is common if male birth and likewise said, mortality of male births is characteristic to firstborns; it suggests that recalling or miss-reporting bias based on gender cannot be excluded. Mortality of women is statistically lower among women in male firstborn families suggesting selection on women from female firstborn families. This is consistent with the stopping rule theory where women from female firstborn families are likely to be part of bigger families (albeit not statistically different) with scarcer resources to share among siblings. I discuss and take into account these differences in the Empirical Strategy section.

Tab. 21: Mean Difference by Gender of Firstborn and Survival Status

Variable	Obs	Mean				Obs	Mean			
		Female FB	Male FB	Diff	Pval		Female FB	Male FB	Diff	Pval
age	35730	27.5365	27.5563	0.02		28461	27.1223	26.9293	-0.193	*
years of education	35693	5.3818	5.2856	-0.096		28428	5.5625	5.4384	-0.1241	*
head is male	35730	0.8041	0.8107	0.007		28461	0.8024	0.815	0.0126	**
poorer	35730	0.1945	0.2029	0.008		28461	0.187	0.1988	0.0118	**
poorest	35730	0.1866	0.1846	-0.002		28461	0.184	0.1815	-0.0025	
middle	35730	0.1951	0.1946	-0.001		28461	0.1927	0.1943	-0.0016	
richer	35730	0.2017	0.1996	-0.002		28461	0.2053	0.2024	-0.0029	
richest	35730	0.2221	0.2183	-0.004		28461	0.231	0.2229	-0.008	
household members	35730	6.7708	6.7347	-0.036		28461	6.6879	6.6658	-0.022	
respondent is head	35730	0.0947	0.0946	0		28461	0.0925	0.0886	-0.0039	
sons died	35730	0.2595	0.2603	0.001		28461	0.2361	0.2365	-0.0004	
dauthers died	35730	0.2304	0.2227	-0.008		28461	0.2131	0.1995	-0.0136	*
total siblings	35730	6.0791	6.0509	-0.028		28461	5.9198	5.8197	-0.1002	***
siblings's children	35730	1.9173	1.2555	-0.662	***	28461	0.7689	0.9516	0.1827	
first born dead	30442	0.1854	0.2188	-0.026	***					
under five died	30442	0.6994	0.7365	0.0371	**	28461	0.4951	0.4557	-0.0394	***
siblings died	35730	1.1481	1.2025	0.0544	***	28461	0.7852	0.7396	-0.0456	***
female siblings died	35730	0.632	0.4521	-0.18	***	28461	0.3364	0.3223	-0.0141	

Note: Sample consists of all non firstborn women aged between 15-49 with at least one sibling. Standard errors are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.10

Empirical Strategy

I take advantage of a quasi-random variation on gender of firstborns to test the role of older brothers on risky sexual behavior of teens in patriarchal societies. I measure risky sexual behavior of teens by out-of-wedlock pregnancy given self reported variables on sexual behavior can be subject to bias. I compute premarital live births of representative women aged between 15-49 from Central Africa. The identification strategy lies on the idea that gender of the firstborn among siblings is exogenous to household characteristics, under the assumption that selective abortion is inaccessible to these women and almost all women have at least one child.[39] In Africa, exogeneity of sex of the firstborn is likely to be at birth but, in subsequent years, women might have followed different patterns in life that can be attributable to gender of the firstborn in their family. I have discussed some of these patterns and motivations in the previous paragraph: *firstly*, "son preference" where women tend to grow with

[39] To my knowledge, there is no prenatal sex detection technologies in Sub-Saharan Africa that can put in cause the validity of the inference like in Asia (Arnold *et al.*, 2002)

higher number of siblings; *secondly*: miss-reporting/recalling bias where older and less educated women tend to declare being from firstborn family; *third*, the correlation between nutritional status of the mother and sex of the child might further put in cause the validity of causal inference. Indeed, sex preference induce selection as mortality among women from poorer and female firstborn families is likely to be high; miss-reporting/recalling bias would also suggest that women with high rank among siblings (further from the firstborn) are likely to be from male firstborns. Precisely, women from bigger and poorer socio-economic condition are likely to miss-report being from male firstborn families. The correlation between nutritional status of the mother and sex of the child would also imply high mortality among female firstborn families. In conclusion, women from female firstborn families sampled are likely to be better off in terms of socio-economic condition inducing a downward bias on the role of male firstborn families on teen pregnancy. To account for this bias, I control for the following: general socio economic conditions: regional, age, wealth, religion, total number of siblings and rank fixed effects. Other controls are education, household size, age and sex of the head of household and the variables that are statistically different between the two groups as shown in Table 21.

I first estimate regression of the following form:

$$y_{ih} = \alpha + \beta male_{ih} + X'_{ih}\gamma + \lambda_r + \varepsilon_{ih} \tag{9}$$

where y_{ih} is premarital pregnancy encountered by woman i in household h and region r, $male$ takes the value one if the teen is born in a household with male firstborn based on birth history of the biological mother and zero else. In some specifications, male takes ordinal values that takes into account the number of preceding males. X_{ih} represent household characteristics and $\lambda_r$are regional dummies. X_{ih} include age dummies, education, age of the head of the household, if female headed household, birth interval and dummies for income of the household. In a first stage, I restrict the sample to secondborns in order to avoid selection due to son preference and miss-recalling/reporting. Conditional on a girl being secondborn, gender of the firstborn allows a causal interpretation. As gender of the firstborn alter the number of siblings a woman has, in all regressions I control for dummy that accounts for the total number of siblings. I use fixed effects for the number of preceding siblings to control for miss-recalling and son preference of the mother. Those in higher ranks are likely to be from female firstborn families if the stopping rule applies for the mother. Age and

region fixed effects control for all other factors that are specific to cohorts and the region and are likely to determine family structures and behavior of women. For ease of interpretation and analysis, I omit families with twin births. Intracluster correlation within households might alter standard errors, hence all regressions are clustered at household level.

I further conduct survival analysis to model if the timing of age at first premarital birth depends on gender of the first born from the mother's family. To keep it in comparable terms, I run the cox hazard regression on secondborns. I estimate the following model:

$$h_i(t) = h_0(t)exp(\beta_1 x_{i1} + .. + \beta_k x_{ik}) \tag{10}$$

where $h_0(t)$ is the baseline hazard at time t and X_{ik} are the covariates or risk factors that determine premarital pregnancy, including the variable of interest that is male firstborn families.

If gender roles and age-based hierarchy is typical in Central Africa, one would expect the parameters on male firstborns to be negative and to increase with the increasing number of preceding males. It is likely that hierarchy might be based both on gender and age. If age based hierarchy among siblings is predominant, the role of an older brother is likely to be different from that of a younger brother in siblings relationships. Gender of a younger sibling is exogenous to behavioral choices of the mother given she has conscious decisions over the number of her births; gender of the first child is likely to shape the behavior of the mother if having an additional child or not. Conditional on deciding to have another child, sex of the younger sibling is exogenous to behavioral choices of the mother. I control for socio-economic characteristics that are likely to differentiate women with a younger brother from those with sisters in their subsequent life. I substitute gender of the next younger brother in the above Equation 9. If gender roles matter more in shaping siblings behavior, the coefficient should be negative and significant as that of the older brother.

4.3 Results

The Effect of Male Firstborn on Out-of-Wedlock Teen Pregnancy

In the sample of teenage girls, more than 97.5% have at least one sibling; this allows a causal interpretation on the effect of gender of firstborns on secondborn's behavior. Table 22 gives OLS estimates of male firstborns on teen pregnancy. The sample refers

to secondborn teens between the age of 15-19.[40] My analysis is based on never married women and their pregnancy.[41] I first look at the youngest cohort and secondborns because selection is likely to be minimized in such a sample. Indeed, from Column (1)-(7) the coefficient on male firstborn is stable over the inclusion of covariates that takes into account household characteristics. Relative to the average pregnancy rate of their counterparts in female firstborn, i.e. 0.05, being born in a male firstborn family reduces out-of-wedlock pregnancy by .025 points. This corresponds to a fifty percent decrease in relative terms. The stability of the coefficient across all columns suggest exogeneity of parental characteristics on secondborns.[42] From descriptive statistics, women of older ages and higher ranks are more likely to be from male firstborn families and their characteristic differs across gender of the firstborn. In regressions with the overall sample, we include household characteristics that take into account this selection. In Column (5), I include a dummy if the household is female headed; the presence of the father in the household reduces teen pregnancy like the male first. The magnitude of the coefficient is similar to that of an older brother.

Survival analysis on the incidence of child birth reveals that having an older brother accelerates the hazard of out-of-wedlock childbearing. In Figure 11 I plot the hazard ratio by gender of the firstborn. Male firstborn reduces the incidence of premarital pregnancy. Table 23 puts magnitude on the previous graph with results from the cox hazard regression. Being secondborn in male firstborn families reduce the hazard ratio by 34%.

Results from the above two regressions give evidence of the role older brothers have in reducing teen pregnancy and the associated costs. This effect is less likely to be correlated with household characteristics or parental preferences.

[40] Results based on a sample of women from 15-24 give very similar and stronger results. As discussed above, if individuals are subject to miss-recalling/reporting, then sex of firstborns might be correlated with other characteristics. I focus on girls up-to the age of twenty-four, which are the youngest cohorts of the sample and potentially more capable of reporting an accurate information on older sibling's sex composition.

[41] Definition of marriage and average age at first marriage change by region, culture and religion. This might alter the definition of woman and transition to adulthood. The role of brothers as watchdogs on sisters not only can prevent premarital pregnancy but also induce gun-shot marriage. I consider the extreme case where the women never entered marriage. This would create a downward bias as we do not consider women who had premarital pregnancy and were forced into marriage.

[42] The coefficient varies a lot across countries with the biggest magnitude in Cameroon. The coefficients in Nigeria and Chad are negative but not statistically significant because average teen pregnancy is very low.

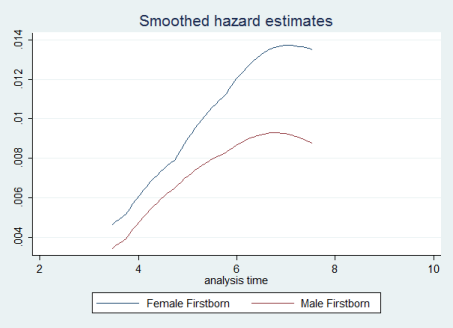
Tab. 22: Male Firstborn Effect on Teen Pregnancy of Never Married Teen Secondborns

	Premarital Childbeaing						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male Firstborn	-0.0269***	-0.0258***	-0.0254**	-0.0255**	-0.0246**	-0.0245**	-0.0247**
	(0.0100)	(0.00991)	(0.01000)	(0.00999)	(0.00995)	(0.00996)	(0.01000)
HH Head Education				-0.000240	-0.000190	-0.000219	-0.000243
				(0.000227)	(0.000233)	(0.000224)	(0.000224)
HH Head Sex					0.0253*	0.0244*	0.0238*
					(0.0141)	(0.0143)	(0.0143)
Years of Education						0.00118	0.00125
						(0.00137)	(0.00137)
URBAN							-0.0163
							(0.0152)
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.0519***	0.0858***	0.0813***	0.0822***	0.0764***	0.0726***	0.0733***
	(0.00851)	(0.0173)	(0.0213)	(0.0213)	(0.0214)	(0.0212)	(0.0213)
Observations	1,618	1,618	1,618	1,618	1,618	1,617	1,617
R-squared	0.062	0.073	0.073	0.073	0.076	0.076	0.077

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Sample: Secondborn Teenage Girls in Singleton Families.

Notes: OLS estimates of male firstborn on teen pregnancy on the sample of secondborn females aged between 15-19. The standard errors are clustered at household level. The sample includes surviving women with at least one ever born older sibling and only observations residing in households without twin births are included. All regressions include regional fixed effects and age fixed effects.

Fig. 11: Cox Hazard Estimates on Secondborn Women aged between 15-24



Tab. 23: Cox Hazard Estimates on Age at First Birth

Cox Proportional Hazard	
Male Firstborn	0.667**
	(0.108)
Female Headed Household	1.684***
	(0.308)
Years of Education	1.022
	(0.0144)
URBAN	0.568***
	(0.107)
Birth Interval	1.074***
	(0.0267)
Observations	3,045

Robust seeform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes:The sample includes women aged between 15-24 who are secondborns in singleton birth families. The standard errors are clustered at household level and the coefficients are the proportional hazard ratios. Controls include age, ever had sexual relation and religion. The regression is based on the pooled data of Cameroon, Nigeria and Chad.

4.4 Mechanisms and Alternative Explanations

The Role of Gender and Birth Order. Older Brothers as "watch-dogs" Older male siblings, other than firstborns, are likely to have similar effects on premarital pregnancy of younger siblings. As discussed in the descriptive statistics, the identification strategy for male older siblings might however be put in cause by selection on parental preferences and characteristics. Table 24 shows the impact of preceding number and proportion of males on teen premarital pregnancy. The first two columns in Table 24 show that having one additional male preceding brother would

reduce the likelihood of premarital pregnancy by 1 percentage points. When I control for socio-economic characteristics, the coefficient increases confirming the downward bias discussed in the empirical strategy. Alternatively, Column (3) and (4) illustrate that moving from no preceding male to all males would reduce the propensity to pregnancy by 3 percentage points. Both the parameters on the number of preceding males and its proportion are comparable to the coefficient found on male firstborns in the above regressions.

There are two theories through which older siblings influence the younger ones: the role and opportunity model. The role model states that younger siblings tend to imitate older ones while in the second mechanism, older siblings influence younger ones by providing opportunities (friends and settings) which might include substance use and sexual intercourse.[43] This paper highlights the second mechanism with the effect being gender related. Female firstborns can also exercise a direct control over their younger teens if they communicate about sexuality with each other. This is not the case in Central Africa and teen girls seem to communicate more with friends rather than family members about topics related to contraception use. In either case, this would not affect the conclusion of the study.

In Column (5) and (6) of Table 24, I estimate the impact of gender of the next younger sibling on premarital pregnancy. Given a teen has a younger sibling, gender of the next younger sibling can be interpreted as random. Results suggest that premarital pregnancy is not affected by a sex of younger sibling. This would suggest that age-based hierarchy predominates over gender roles among siblings. Precisely, age-based hierarchy interacted with gender is the main channel that drives the above results.[44]

Columns (7) and (8) of Table 24 show the role of preceding males in female firstborn families. The coefficients on preceding males is not significant suggesting the importance of male firstborn brothers in influencing risky sexual behavior of younger sisters and acting as watchdogs in their transition to adulthood. Reporting a citation, discussed in Section 4.1, it is likely that male firstborns are the most dominant figure among siblings. *In the context of intra-household relationship, the ruler is the eldest*

[43] Our results might be driven by the role model if women from female firstborn families are likely to have older sisters with premarital pregnancy. Unfortunately we do not have data on birth history of siblings to identify any causal inference of older siblings behavior on younger ones. The impact of preceding males might also be driven by the opportunity model if sister are likely to discuss family planning and share information.

[44] One reason might be age distance between older siblings and potential partners of younger sister. The smaller the age gap between the brother and partner of the sister, the better it is for the brother to screen and retrieve information on the quality of partners.

Tab. 24: Preceding and Succeeding Males on Premarital Pregnancy

	Dependent Variable is Premarital Childbearing									
	All Non-Firstborn Families				All Non-Lastborn Families		Female Firstborn Families		Third Born Women	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Preceding Num of Males	-0.00882**	-0.00929**					-0.0116			
	(0.00449)	(0.00440)					(0.00769)			
Preceding Proportion of Males			-0.0307**	-0.0314**				-0.0456		
			(0.0139)	(0.0138)				(0.0304)		
Post-Male					-0.00624	-0.00609				
					(0.00844)	(0.00827)				
Male Male									-0.0780***	-0.0702**
									(0.0283)	(0.0278)
Female Female									-0.00354	0.00391
									(0.0284)	(0.0284)
Years of Education		-0.0208***		-0.0201***		-0.0201***	-0.0215***	-0.0215***		-0.0232***
		(0.00221)		(0.00236)		(0.00200)	(0.00326)	(0.00326)		(0.00422)
Male headed HH		-0.0257**		-0.0192*		-0.0247**	-0.0209	-0.0211		0.00357
		(0.0105)		(0.0103)		(0.0106)	(0.0153)	(0.0153)		(0.0230)
Ever used Contraception		0.119***		0.108***		0.119***	0.128***	0.128***		0.106***
		(0.0151)		(0.0147)		(0.0152)	(0.0225)	(0.0225)		(0.0324)
Mortality among Siblings		0.00946*		0.0109*		0.0116**	0.0100	0.00997		0.0341**
		(0.00528)		(0.00604)		(0.00521)	(0.00752)	(0.00753)		(0.0160)
URBAN		-0.0129		-0.0107		-0.0126	-0.00118	-0.000783	0.000906	0.000762
		(0.0126)		(0.0125)		(0.0124)	(0.0178)	(0.0178)	(0.0268)	(0.0267)
Constant	0.658**	0.732***	0.701***	0.773***	1.056	1.079	0.514	0.477	0.368	0.283
	(0.262)	(0.258)	(0.250)	(0.250)	(0.683)	(0.717)	(0.361)	(0.354)	(0.302)	(0.298)
Observations	8,968	8,954	8,968	8,954	9,629	9,615	4,148	4,148	1,739	1,739
R-squared	0.189	0.214	0.187	0.210	0.193	0.219	0.245	0.244	0.323	0.345

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes:OLS estimate of the number and proportion of preceding males from Column(1)-(4). In Column (5) and (6) I show estimates for younger succeeding brother while in Column (7) and (8), I restrict the sample to female firstborn families and evaluate preceding number of males. In the last two columns, I look at third born girls and sex composition of the first two siblings. Only non firstborn teenage girls in families of singleton births are included in the sample. I omit rank when the sample is specific to women of a given rank. Standard errors are clustered at the household level. All regressions include regional, age, rank. religious, total number of siblings and income fixed effects.

son, he dominates the younger siblings. Elder and younger sibling's relationship exist only among males; women are seen as instruments of domination by firstborn (Marc-Eric Gruénais, 1985). However, a further explanation would be statistical power as variability in the dependent variable is very low. Indeed the coefficient remains negative. Selection on the number of siblings and systematic difference between the two groups, as discussed in the empirical strategy, might be another reason for the downward bias.

The last two columns evaluate sex composition of preceding siblings on thirdborns. Third born women with two preceding males are less likely to have premarital pregnancy as compared to mixed preceding siblings confirming the above results. Indeed the magnitude of the coefficient is almost three times the parameter on secondborns. The role of male firstborn is the most important determinant of premarital pregnancy as in the previous regressions. This suggests that gender of the firstborn makes the difference in shaping risky sexual behavior of teens. Preceding males are likely to have similar effects but endogeneity constrains the causal impact on premarital pregnancy.

Interaction among firstborns and younger siblings: survival status of firstborns, age distance and presence of male figure in the Household I test different mechanisms to argue that results support the idea of male brothers as "*watchdog*s" to shape the behavior of younger teens. Precisely, I show that direct interaction between firstborns and non is necessary to shape premarital pregnancy. For example, male births are likely to die in early ages and this would imply that mortality of male firstborns is higher. I test if survival of firstborn is important in capturing the impact of male firstborns on premarital pregnancy. If it is the case, this would imply that interaction between siblings is essential in reducing premarital pregnancy. In Table 25, I show coefficients on survival status of the firstborn. Column (1) shows the coefficient on male firstborns alive at the time of the survey, or have lived at least until the age of 20, with a negative and significant coefficient. Alive male firstborn reduces the likelihood of pregnancy by .018, which is comparable with the above coefficient on second born teenage girls. Women born in families with male firstborn who died before the age of 20 are likely to be pregnant out of wedlock. The interaction coefficient is positive and big in magnitude.[45] This is consistent with the idea that survival

[45] It might be the case that families with male firstborn who died before the age of 20 are worse-off and the coefficient might be capturing this economic condition. Nonetheless, these women are doing worse than women born female firstborns who died before the age of 20. As discussed in the descriptive statistics, there might be selection on gender of the firstborn and these women might be doing better with

of firstborns is a determinant element to impact negatively premarital pregnancy and sexual behavior of younger siblings.

I next test if distance from the firstborn matters for premarital pregnancy. The assumption is that more distance would imply that the firstborn is farther from networks of the younger siblings and would thus reduce monitoring and supervising. In Table 25, I show that distance from firstborns decreases the capacity of supervision and monitoring. It is intuitive as distance increases the likelihood of siblings to belong to different networks. Moreover, if distance of the firstborn is too high as compared to the teen, the likelihood of both siblings to reside in the same household decreases. I measured distance as the difference between the age of the firstborn and that of the woman. The coefficient on gender of the firstborn is negative while the interaction term is positive suggesting that the negative impact of male firstborn on premarital pregnancy decreases as distance increases. If age distance between the male firstborn and younger sister is one, the role of "watchdog" by the brother would decrease premarital pregnancy by 3.1% points. If distance between the two is 5 years then the reduction in premarital pregnancy would be by 2% points reaching almost no effect if distance is above 15 years. The results confirm that male firstborns who are closer to networks of younger sisters are more effective in monitoring and supervising their behavior.

If male firstborns serve as watchdogs, it is presumable to think that male headed families or biological fathers have similar effects on pregnancy of teens. The interaction based on the presence of both in the household should also presumably further reduce premarital pregnancy. In Column (3) of Table 25, I test this last mechanism. The coefficients on male headed families, male firstborns and their interactions are not significant with pvalues around 0.11-0.13. The sample is restricted to women classified as daughters of the head of household and the variance of the dependent variable is further reduced. I discuss the parametres assuming that the low pvalue is due to the reduction in the variability of premarital pregnancy. Women in male headed families with male firstborn families are likely to avoid premarital pregnancy with magnitude that is higher than the 2.5 percentage points in the above analysis. This suggests that the presence of the father in the household further reduces premarital pregnancy.

---

respect to male firstborn families.

Tab. 25: Mechanisms and Siblings Interaction on Premarital Pregnancy

	Dependent Variable is Premarital Pregnance		
	Survival	**Age Distance**	**Male Headed HH**
	(1)	(2)	(4)
Male Firstborn	-0.0176*	-0.0332**	0.0345
	(0.00924)	(0.0163)	(0.0226)
Firstborn died before 20	-0.0171		
	(0.0230)		
Male Firstborn*Firstborn died before 20	0.0618**		
	(0.0313)		
Male Firstborn*Age distance from FB		0.00248*	
		(0.00148)	
Density			
Male Firstborn*Density			
Male Firstborn*Male headed HH			-0.0374
			(0.0245)
Male headed HH	-0.0414***	-0.0407***	-0.0281
	(0.0115)	(0.0112)	(0.0187)
Age distance from FB		0.00590**	
		(0.00261)	
Age distance from FB^2		-0.000116**	
		(5.07e-05)	
Years of Education	-0.0203***	-0.0203***	-0.0129***
	(0.00217)	(0.00198)	(0.00208)
HH Members	0.00695***	0.00708***	0.00626***
	(0.00146)	(0.00135)	(0.00162)
Ever used Contraception	0.119***	0.118***	0.116***
	(0.0150)	(0.0143)	(0.0181)
Mortality among Siblings	0.00899*	0.0109**	0.0180***
	(0.00537)	(0.00523)	(0.00549)
Constant	0.653**	0.568**	-0.121
	(0.257)	(0.263)	(0.0852)
Observations	8,954	8,954	5,730
R-squared	0.217	0.218	0.188

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes:OLS estimate of survival status of male firstborns, age distance and presence of the father in the household. Column(1) shows estimates for male firstborn who had live at least until the age of 20; Column (2) shows how distance from the firstborn reduces the parameter on male firstborns and finally Column (3) shows how fathers have similar effects on teen pregnancy. The last column is based on the sample of women classified as daughter of the head of household. Only non firstborn teenage girls in families of singleton births are included in the sample. Standard errors are clustered at the household level. All regressions include regional, age, rank. religious, total number of siblings and income fixed effects. Other controls are are shown in the table.

To sum up, in SSA, age gap between teenage girls and their partners is relatively high as compared to many industrialized countries. [46] Potential partners of female siblings are closer in age to older brothers rather than younger ones. This gives comparative advantage for an older brother as a "*watchdog*" . The fact that networks are gender biased, older brothers are more advantaged as compared to older sisters in acquiring information on pretender's quality. I give evidence of this mechanism and interaction among siblings in Table 26. It shows that male firstborns increase age gap between

[46] I refer to the extensive literature on inter-generational sex in SSA, emphasized in Dupas (2011).

the woman and her first husband by almost a year. This effect increases to more than a year if brother is alive and age gap with firstborn is low. As expected, having a younger male brother does not affect the characteristics of the partner. It leads to the conclusion that older sibling's gender affects age gap between a woman and her partner at first marriage. Women are likely to choose partners who are older and outside own network to avoid being caught by watchdogs of the family. These results are particularly true for women who got married between the age of 15 and 30 in urban areas.

Tab. 26: The Role of Preceding Males and Age Gap with Partner

	Dependent Variable: Age Gap with Partner			
First Union and Age at Marriage between 15-30 in Urban Areas				
	(1)	(2)	(3)	(4)
Male Firstborn	0.706*		0.908**	1.067**
	(0.409)		(0.424)	(0.476)
Post Male		0.0588		
		(0.406)		
Firstborn died before 20			0.533	
			(1.248)	
Male Firstborn*Firstborn died before 20			-1.846	
			(1.489)	
Male Firstborn*Age distance from FB				-0.105
				(0.0778)
Age distance from FB				0.158
				(0.135)
Age distance from FB^2				-0.000424
				(0.00163)
Male Firstborn*Male headed HH				
Years of Education	-0.265***	-0.246***	-0.265***	-0.264***
	(0.0487)	(0.0516)	(0.0487)	(0.0491)
Male headed HH	0.663	0.617	0.633	0.706
	(0.558)	(0.591)	(0.554)	(0.559)
Mortality among Siblings	0.141	0.104	0.196	0.136
	(0.260)	(0.262)	(0.275)	(0.260)
Constant	8.887**	10.60***	8.944**	8.345**
	(3.682)	(3.691)	(3.860)	(3.686)
Observations	1,165	1,069	1,165	1,165
R-squared	0.112	0.120	0.114	0.115

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes:OLS estimate of age gap with first husband or cohabitant of women in union only once. The sample includes women who got married between the age of 15 and 30 in urban areas. Such a sample excludes forced marriage or parental arrangements with high population density where women are likely to have several contenders. Only non firstborn teenage girls in families of singleton births are included in the sample. Standard errors are clustered at the household level. All regressions include regional, age, rank. religious, total number of siblings and income fixed effects.

Other Mechanisms: Differential allocation of parental resources, siblings sex composition and resource competition would imply that parents allocate less resources to teens with an older brother and hence increase the propensity of pregnancy. This would not affect the main results. One can still assume that parents decide to push the teen girl with an older brother into marriage. In the data-set, there is no evidence of an association between gender of an older brother and the probability of being in marriage while teen nor anticipates the timing of marriage.

In general, inferring on teen pregnancy is challenging due to under-reporting. For-example Leibowitz *et al*. (1986) show that teens with greater value of time are less likely to carry their pregnancy to term, while those who already dropped out of school are more likely to bear the child. In a survey conducted in Cameroon on 384 adolescents aged between 20-29, being in school increased the likelihood that a pregnancy ended in abortion more than 7 times. The results found in previous sections might be capturing the effect of male firstborn on abortion rather than risky behavior. It makes sense to think that women with an older brother are likely to abort rather than teens with an older sister. Abortion is illegal and unsafely practiced in the countries under study and if it is the case, mortality of women should be higher in male firstborn families. This is not the case in the data. If anything else, male firstborns reduce significantly mortality of younger sisters in older ages as shown in Table 21.

Sex of preceding birth significantly affects birth-weight of successive pregnancies. After a male birth, the body of the mother is likely to be worn because it is costly in terms of energy and breast-feeding. One can assume that a teen born after a male firstborn is physically doing worse than a teen after a female birth; this makes the latter more active in the matching and marriage market. The results on male firstborn are not driven by the difference in the Body Mass Index of the teens.

Finally, a leading alternative explanation for the results is the role model, *i.e.* younger siblings tend to imitate older siblings. A secondborn with an older sister is more prone to be pregnant than with an older brother if the sister has been ever pregnant. Older siblings might introduce the younger ones to deviating behaviors and expose them early to adult behaviors. Unfortunately, there are no data available on birth history of siblings to investigate this mechanism. However, the effect might go either way if pregnancy of an older sister alert parents who put more effort in supervising the younger ones. Moreover, many studies who addressed this phenomenon suffer from a drawback on distinguishing between neighborhood or parental effect from that of role model.

4.5 Welfare Consequences of Premarital Pregnancy while Teen

From descriptive statistics, a good percentage of women tend to get married before reaching the age of 19, as shown in Panel F of Table 20. In this section, I run regressions to look at the role of premarital pregnancy when teen on marital status, quality of marriage and the child's well-being. Given the average age at first birth is sixteen, I further distinguish between premarital pregnancy while teen before and after sixteen[47]. The excluded categories are those who got pregnant after marriage or never got pregnant while teen.

Premarital Pregnancy and Marital Status

Table 27 reports results on marital status of women who had premarital teen pregnancy, subdivided by age-group. Never married women in their twenties, at the time of the survey, are associated with being head of the household if they had out-of-wedlock pregnancy. This holds specially for women who got pregnant before the age of sixteen. Being head of the household might be costly in a pro-male biased economy, although difficult to conclude on the overall welfare.

Premarital pregnancy reduces the likelihood of being ever married among older cohorts and for those pregnant after the age of sixteen. The youngest cohorts, between the age of 20-29 with out-of-wedlock pregnancy before the age of sixteen, are more likely to be married. Among those ever married, Panel B shows they are also more likely to be classified as wife of the head of household. On the reverse, teens pregnant out-of-wedlock in older cohorts or after the average age of sixteen have different patterns as compared to those pregnant before sixteen in the youngest cohort. As what concerns age at marriage, as expected, premarital pregnancy delays age at marriage. The delay in age at marriage of the youngest cohort seems very small as compared to the older ones, suggesting they enter the marriage market easily.

In conclusion, for the youngest cohorts, age at premarital pregnancy matters for their marital status in older ages. Those pregnant out-of-wedlock before sixteen are more likely to enter the marriage market, be classified wife of the head and get married on average less than one year older compared to their peers without premarital pregnancy. Those who have premarital child *after* the age of sixteen are less likely to get married

[47] The basic idea behind is to disentangle the role of pregnancy after the age of sixteen, which might be considered the norm. Among those who had a pregnancy before the age of nineteen, sixty percent of them were pregnant before the age of sixteen.

and enter the marriage market much more later than their similar.

For the oldest cohorts, independently from age of pregnancy, they are less likely to ever get married. Moreover, among those ever married, the delay of age at marriage is higher as compared to their cohorts.

Premarital Pregnancy and Quality of Partner

Table 28 shows the association between premarital teen pregnancy and partner's quality for ever married women. Ever married women in the oldest cohorts seem not to be affected by premarital pregnancy in terms of age-gap with partner and his years of education compared to their cohorts without premarital pregnancy. In the same line, those who had premarital pregnancy after the age of sixteen do not seem to be affected in terms of partner's quality. Conversely, teens who get pregnant out-of-wedlock before the age of sixteen and belong to the cohorts of 20-29 or 30-39 are more likely to be married with a less educated partner and older one.

In conclusion, for the youngest cohorts, age at premarital pregnancy matters for the quality of their partner. They are more likely to be married to less educated men with higher age-gap compared to their cohorts without premarital pregnancy. Older cohorts and those pregnant after the age of sixteen seem not to be affected by teen premarital pregnancy in terms of partner's age and education.

Premarital Pregnancy and the Child's Well-being

I investigate the survival status and residence of the firstborn child, born out-of-wedlock, compared to those within wedlock. Table 29 shows different patterns among different cohorts regarding the residence of the child. Women of older cohorts, aged between 30-49 and got pregnant out-of-wedlock, are less likely to have their child living with them as compared to their cohorts who ever gave birth within marriage. As for the youngest cohort, where age at pregnancy matters for the well-being of the child, premarital pregnancy before the age of 16 is associated with a negative likelihood of the child living with the mother while the opposite holds for those pregnant after the age of 16.

Across all cohorts, having a premarital child while teen is associated with the likelihood of the child being dead. These is in line with the literature on the risks of neonatal

Tab. 27: Premarital Pregnancy and Marital Status across Cohorts

Welfare Implication of Out-of-Wedlock Teen Pregnancy on Marital Status

Column	I	II	III	IV	V	VI	VII	VIII	IX	
AGE	Panel A: Probability of being Head of Household									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	0.0402*	0.0557**	0.0576**	0.118**	0.135***	0.138***	0.00897	0.0233	0.0246	Never Married
	(0.0234)	(0.0235)	(0.0236)	(0.0468)	(0.0468)	(0.0465)	(0.0254)	(0.0255)	(0.0256)	
	Probability of being Ever Married									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	-0.00291	-0.0291**	-0.0288**	0.0626***	0.0260**	0.0265**	-0.0701***	-0.0829***	-0.0829***	
	(0.0119)	(0.0114)	(0.0114)	(0.0136)	(0.0132)	(0.0132)	(0.0186)	(0.0179)	(0.0179)	
30-39	-0.0113	-0.0150**	-0.0149**	0.00230	-0.00301	-0.00292	-0.0281**	-0.0288**	-0.0287**	All
	(0.00721)	(0.00720)	(0.00721)	(0.00755)	(0.00757)	(0.00757)	(0.0128)	(0.0128)	(0.0128)	
40-49	-0.0228***	-0.0205***	-0.0203***	-0.0136*	-0.0125*	-0.0122*	-0.0344***	-0.0307**	-0.0306**	
	(0.00712)	(0.00681)	(0.00684)	(0.00780)	(0.00733)	(0.00737)	(0.0131)	(0.0127)	(0.0127)	
	Panel B: Probability of being Wife of the Head of Household									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	0.0148	0.00775	0.0103	0.0536***	0.0446**	0.0470***	-0.0364	-0.0398*	-0.0370	
	(0.0147)	(0.0147)	(0.0147)	(0.0173)	(0.0173)	(0.0174)	(0.0243)	(0.0242)	(0.0242)	
30-39	0.00907	0.00610	0.00815	-0.00398	-0.00881	-0.00719	0.0227	0.0228	0.0254	Ever Married
	(0.0148)	(0.0148)	(0.0148)	(0.0180)	(0.0180)	(0.0180)	(0.0238)	(0.0238)	(0.0238)	
40-49	-0.0556***	-0.0562***	-0.0536***	-0.0408*	-0.0418*	-0.0380*	-0.0769**	-0.0769**	-0.0760**	
	(0.0190)	(0.0190)	(0.0190)	(0.0230)	(0.0230)	(0.0229)	(0.0314)	(0.0314)	(0.0313)	
	Panel C: Age at First Marriage									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	1.498***	1.648***	1.670***	0.369***	0.558***	0.582***	2.906***	2.979***	2.995***	
	(0.0967)	(0.0959)	(0.0951)	(0.117)	(0.116)	(0.115)	(0.123)	(0.126)	(0.125)	
30-39	2.496***	2.656***	2.673***	1.040***	1.273***	1.286***	4.274***	4.296***	4.320***	Ever Married
	(0.161)	(0.162)	(0.162)	(0.191)	(0.192)	(0.192)	(0.238)	(0.248)	(0.249)	
40-49	3.780***	3.904***	3.896***	2.547***	2.791***	2.770***	5.325***	5.263***	5.273***	
	(0.268)	(0.265)	(0.265)	(0.342)	(0.339)	(0.339)	(0.394)	(0.393)	(0.393)	
Years of Education FE		Yes	Yes		Yes	Yes		Yes	Yes	
Total Siblings FE			Yes			Yes			Yes	
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
INCOME FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
URBAN FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The first three columns are based on premarital pregnancy while teen. From Column 4-6, it is based on premarital pregnancy before the age of 16 and in the last three columns between 17-19. I use all the sample of women in their reproductive age and look at their marital status by sub-groups based on age. In the latter columns, I exclude all women who got pregnancy before the age of 16. As usual, standard errors are clustered at household level.

Tab. 28: Premarital Pregnancy and Marriage Quality across Cohorts

Welfare Implication of Out-of-Wedlock Teen Pregnancy on Quality of Partner

Column	I	II	III	IV	V	VI	VII	VIII	IX	
AGE	Panel A: Partner's Years of Education									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	-0.540***	-0.137	-0.0933	-0.953***	-0.424**	-0.378**	0.0584	0.251	0.281	
	(0.149)	(0.133)	(0.128)	(0.187)	(0.167)	(0.161)	(0.221)	(0.199)	(0.193)	
30-39	-0.382**	-0.122	-0.0938	-0.762***	-0.336*	-0.265	0.242	0.213	0.184	Ever
	(0.174)	(0.147)	(0.144)	(0.216)	(0.179)	(0.174)	(0.259)	(0.226)	(0.222)	Married
40-49	-0.311	-0.0416	0.0294	-0.767***	-0.220	-0.111	0.392	0.232	0.251	
	(0.192)	(0.168)	(0.163)	(0.234)	(0.195)	(0.190)	(0.304)	(0.284)	(0.277)	
	Panel B: Partner's Age									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	0.686***	0.557**	0.583**	1.093***	0.930***	0.954***	0.0933	0.0269	0.0543	
	(0.253)	(0.254)	(0.254)	(0.322)	(0.323)	(0.323)	(0.370)	(0.371)	(0.370)	
30-39	0.852***	0.662**	0.652**	1.454***	1.174***	1.171***	-0.145	-0.172	-0.185	Ever
	(0.310)	(0.308)	(0.308)	(0.405)	(0.400)	(0.400)	(0.435)	(0.437)	(0.437)	Married
40-49	-0.0858	-0.254	-0.248	0.399	0.112	0.135	-0.814	-0.790	-0.805	
	(0.382)	(0.379)	(0.378)	(0.475)	(0.473)	(0.472)	(0.602)	(0.597)	(0.596)	
Total Siblings FE			Yes			Yes			Yes	
Years of Education FE		Yes	Yes		Yes	Yes		Yes	Yes	
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
INCOME FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
URBAN FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The first three columns are based on premarital pregnancy while teen. From Column 4-6, it is based on premarital pregnancy before the age of 16 and in the last three columns between 17-19. I use all the sample of women in their reproductive age and look at their marital status by sub-groups based on age. In the latter columns, I exclude all women who got pregnancy before the age of 16. As usual, standard errors are clustered at household level.

morbidity and mortality associated with teen pregnancy[48]. In fact, the coefficients on pregnancies before 16 are as twice as those between 17-19.

Concluding remarks on Welfare Consequences of Premarital Pregnancy

In countries where age at first marriage and birth are concentrated during teenage, the transition from childhood to adulthood is very early making it difficult to differentiate between premature accidental pregnancy and the norm. To this purpose, the above analysis is distinguished by pregnancies that occurred below and above the average age of birth when teen, i.e. 16. A further distinction is made by age groups.

Family structure and marriage characteristics are changing across generations and cohorts and the results give evidence of this by looking at different cohorts and the welfare consequences due to teen pregnancy.

Women in their twenties at the time of the survey, the welfare consequences of premarital pregnancy before the age of sixteen is completely different compared to those who gave birth within marriage or in older ages. They are more likely to end up in marriage with their child being fostered or dead. Their partners are less educated and older as compared to partners of their cohorts. On the reverse, if they are never married, they tend to be head of the household and live on their own. Overall, having a premarital child before the age of sixteen seems to be associated with marriage, of bad type with non-optimal consequences for the child.

For the oldest cohort, having a premarital pregnancy when teen tends to reduce the likelihood of marriage and delay age at marriage. Nevertheless, the quality of partner among those married, in terms of age and education does not seem to be significantly different from their cohorts. Child's survival and residence have a very similar pattern across cohorts. Children born from premarital pregnancy are more likely to be fostered and less likely to be alive at the time of the survey.

4.6 Conclusion

In Central Africa, age at marriage is increasing due to schooling of girls, leading to the exposure of teens to premarital sex and hence pregnancy. Still, age at first birth within and out-of-wedlock is concentrated during teen-age with an important implication on well-being of the girl. This paper studies how preceding sex composition of

[48] If I restrict the sample to teens with birth before the age of 16, there is no statistical difference in terms of survival if within or outside-wedlock

Tab. 29: Premarital Pregnancy and Survival Status of First Child across Cohorts

Welfare Implication of Out-of-Wedlock Teen Pregnancy on Respondent's Firstborn

Column	I	II	III	IV	V	VI	VII	VIII	IX	
AGE	Panel A: Probability that Firstborn Child Lives with Respondent									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	0.0202	0.00401	0.00326	-0.0337	-0.0572***	-0.0577***	0.0715***	0.0638***	0.0629***	
	(0.0157)	(0.0158)	(0.0158)	(0.0210)	(0.0211)	(0.0211)	(0.0221)	(0.0222)	(0.0222)	
30-39	-0.186***	-0.184***	-0.183***	-0.214***	-0.211***	-0.208***	-0.131***	-0.132***	-0.132***	Ever Birth
	(0.0162)	(0.0163)	(0.0162)	(0.0196)	(0.0196)	(0.0195)	(0.0253)	(0.0253)	(0.0254)	
40-49	-0.126***	-0.120***	-0.118***	-0.117***	-0.105***	-0.101***	-0.132***	-0.136***	-0.135***	
	(0.0148)	(0.0148)	(0.0148)	(0.0185)	(0.0188)	(0.0188)	(0.0221)	(0.0218)	(0.0217)	
	Panel B: Probability that Firstborn of Respondent is Dead									
	Pregnant before 19			Pregnant before 16			Pregnant after 16			
20-29	0.0771***	0.0703***	0.0700***	0.106***	0.0963***	0.0963***	0.0454***	0.0418***	0.0413***	
	(0.0116)	(0.0116)	(0.0116)	(0.0172)	(0.0172)	(0.0172)	(0.0144)	(0.0144)	(0.0144)	
30-39	0.0965***	0.0930***	0.0932***	0.127***	0.122***	0.122***	0.0464**	0.0451**	0.0465**	Ever Birth
	(0.0149)	(0.0150)	(0.0150)	(0.0202)	(0.0203)	(0.0203)	(0.0201)	(0.0202)	(0.0202)	
40-49	0.0756***	0.0736***	0.0732***	0.0872***	0.0822***	0.0815***	0.0544*	0.0569**	0.0568**	
	(0.0190)	(0.0190)	(0.0190)	(0.0248)	(0.0248)	(0.0249)	(0.0278)	(0.0278)	(0.0278)	
Years of Education FE		Yes	Yes		Yes	Yes		Yes	Yes	
Total Siblings FE			Yes			Yes			Yes	
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
INCOME FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
URBAN FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The first three columns are based on premarital pregnancy while teen. From Column 4-6, it is based on premarital pregnancy before the age of 16 and in the last three columns between 17-19. I use all the sample of women in their reproductive age and look at their marital status by sub-groups based on age. In the latter columns, I exclude all women who got pregnancy before the age of 16. As usual, standard errors are clustered at household level.

siblings shape these consequences in strong patriarchal societies. In countries where age-based hierarchy prevails among siblings and gender role is biased towards males, the study finds that being born in a family with male firstborn reduces the likelihood of premarital pregnancy. I used a simple natural experiment on sex of firstborns to investigate the causal impact of gender, interacted with birth-order, on risky behavior of women in Central Africa. I find that male firstborns, who are closer to the network of younger sisters' pretenders, reduce teen pregnancy by 2.5 percentage points, that is equivalent to a 50% reduction. Gender alone does not drive this behavior; indeed, birth order seem to be an important factor that allow older brothers to serve as watch-dogs for pretenders. Additional preceding males in the household serve as watchdogs as well. The mechanism is based on direct interaction among siblings and potential partners. Indeed, distance in age, survival status of the firstborn and presence of the father in the household are other elements that shape the parameter of interest. Brothers tend to socialize more outside the household as compared to older sisters and this puts brothers in a better position to be informed and dismiss in advance, as compared to sisters or younger brothers, siblings' contenders who are of bad type. The study further shows how socio-economic consequences of premarital pregnancy in early ages is heterogeneous across cohorts and age of birth. The youngest cohorts who get pregnant before the age of 16 are likely to have the worst outcome in terms of marriage, husband's quality and survival of the child.

I used data from a population based survey at the individual level from Central Africa to show the role of brothers in patriarchal societies. The study contributes to the existing literature in two manners. Firstly, it extends the findings on the role of birth order and sex composition of siblings on risky behavior of the younger ones with everlasting economic consequences on their well-being. Indeed, previous studies have found that being a non firstborn increases the likelihood of engaging in risky behavior like smoking tobacco, marijuana and sexual behavior (see for example Ouyang 2004; Argys *et al.,* 2006; Averette *et al.,* 2011). This study brings forth the literature by exploiting gender of the firstborn to estimate a causal effect because birth order alone is confounded with total number of siblings. An additional originality of this finding stems from understanding how both birth-order and sex composition of preceding siblings matter in shaping risky behavior of adolescents in patriarchal contexts.

The study also contributes to the existing literature on son preference of women in developing countries. Economic outcomes like inheritence are found as driving forces of the strong preference for sons in developing countries. This study contributes to the

literature in providing evidence of the role sons have in reducing costs associated to women's premarital pregnancy within the household. The rise in age at first marriage has exposed teens to premarital sex and hence potential out-of-wedlock pregnancy rendering girls costly.

Although, average premarital childbearing while teen is more or less constant and comparable across cohorts, there is an increasing number of studies pointing out the rise in abortion over recent years. Understanding factors that drive or shape these patterns is essential for addressing future social policies.

References

[1] M. Abélès and C. Collard. *Age, pouvoir et société en Afrique noire*, volume 14. Presses de l'Université de Montréal, 1985.

[2] G.A. Akerlof. A theory of social custom, of which unemployment may be one consequence. *The quarterly journal of economics*, 94(4):749, 1980.

[3] A. Alesina, R. Baqir, and W. Easterly. Public goods and ethnic divisions*. *Quarterly Journal of Economics*, 114(4):1243–1284, 1999.

[4] A. Alesina, A. Devleeschauwer, W. Easterly, and S. Kurlat. Fractionalization. *Journal of Economic Growth*, 8(2):155–194, 2003.

[5] A. Alesina and E. La Ferrara. Participation in heterogeneous communities. *The quarterly journal of economics*, 115(3):847–904, 2000.

[6] A. Alesina and E. La Ferrara. Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800, 2005.

[7] Douglas Almond and Bhashkar Mazumder. Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal-Applied Economics*, 3(4):56, 2011.

[8] J.G. Altonji, S. Cattan, and I. Ware. Identifying sibling influence on teenage substance use. 2010.

[9] V. Amin. Sibling sex composition and educational outcomes: A review of theory and evidence for the uk. *Labour*, 23(1):67–96, 2009.

[10] J.K. Anarfi. Sexuality, migration and aids in ghana-a socio-behavioral study. *Health Transition Review*, 3(Sup), 1993.

[11] C.J. Anderson and A. Paskeviciute. How ethnic and linguistic heterogeneity influence the prospects for civil society: a comparative study of citizenship behavior. *Journal of Politics*, 68(4):783–802, 2006.

[12] Siwan Anderson. The economics of dowry and brideprice. *The Journal of Economic Perspectives*, 21(4):151–174, 2007.

[13] J.D. Angrist and W.N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3):450–477, 1998.

[14] L.M. Argys, D.I. Rees, S.L. Averett, and B. Witoonchart. Birth order and risky adolescent behavior. *Economic Inquiry*, 44(2):215–233, 2006.

[15] F. Arnold, S. Kishor, and TK Roy. Sex-selective abortions in india. *Population and Development Review*, 28(4):759–785, 2002.

[16] S.L. Averett, L.M. Argys, and D.I. Rees. Older siblings and adolescent risky behavior: does parenting play a role? *Journal of Population Economics*, 24(3):957–978, 2011.

[17] O. Bandiera, I. Barankay, and I. Rasul. Cooperation in collective action*. *Economics of transition*, 13(3):473–498, 2005.

[18] G.S. Becker, J.S. Duesenberry, and B. Okun. An economic analysis of fertility, 1960.

[19] G.S. Becker and H.G. Lewis. Interaction between quantity and quality of children, 1974.

[20] Megan Beckett, Julie Da Vanzo, Narayan Sastry, Constantijn Panis, and Christine Peterson. The quality of retrospective data: an examination of long-term recall in a developing country. *Journal of Human Resources*, pages 593–625, 2001.

[21] S. Bertozzi, N.S. Padian, J. Wegbreit, L.M. DeMaria, B. Feldman, H. Gayle, J. Gold, R. Grant, and M.T. Isbell. Hiv/aids prevention and treatment. *Drug Resistance*, page 25, 2006.

[22] J.M. Blair, D.L. Hanson, J.L. Jones, M.S. Dworkin, et al. Trends in pregnancy rates among women with human immunodeficiency virus. *Obstetrics & Gynecology*, 103(4):663, 2004.

[23] S.S. Bloom, C. Banda, G. Songolo, S. Mulendema, A.E. Cunningham, and J. Boerma. Looking for change in response to the aids epidemic: trends in aids knowledge and sexual behavior in zambia, 1990 through 1998. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 25(1):77, 2000.

[24] Walter Bossert, De Ambrosio Conchita, and Eliana La Ferrara. A generalized index of fractionalization. 2008.

[25] R. Boucekkine, R. Desbordes, and H. Latzer. How do epidemics induce behavioral changes? *Journal of Economic Growth*, 14(3):233–264, 2009.

[26] S. Boyer, F. Marcellin, P. Ongolo-Zogo, S.C. Abega, R. Nantchouang, B. Spire, and J.P. Moatti. Financial barriers to hiv treatment in yaoundé, cameroon: first results of a national cross-sectional survey. *Bulletin of the World Health Organization*, 87(4):279–287, 2009.

[27] A. Brandon and C. Bledsoe. The effects of education and social stratification on marriage and the transition to parenthood in greater freetown, sierra leone. In *Workshop on Nuptiality in sub-Saharan Africa: Current Changes and Impact on Fertility, Paris*, 1988.

[28] M. Brockerhoff and A.E. Biddlecom. Migration, sexual behavior and the risk of hiv in kenya. *International Migration Review*, pages 833–856, 1999.

[29] Kristin F Butcher and Anne Case. The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, 109(3):531–63, August 1994.

[30] J. Byakika-Tusiime, JH Oyugi, WA Tumwikirize, ET Katabira, PN Mugyenyi, and DR Bangsberg. Adherence to hiv antiretroviral therapy in hiv+ ugandan patients purchasing therapy. *International journal of STD & AIDS*, 16(1):38–41, 2005.

[31] A.E. Calvès. Abortion risk and decisionmaking among young people in urban cameroon. *Studies in Family Planning*, pages 249–260, 2002.

[32] A. Case and C. Ardington. The impact of parental death on school outcomes: Longitudinal evidence from south africa. *Demography*, 43(3):401–420, 2006.

[33] A. Case, C. Paxson, and J. Ableidinger. Orphans in africa: parental death, poverty, and school enrollment. *Demography*, 41(3):483–508, 2004.

[34] J. Castilla, J. Del Romero, V. Hernando, B. Marincovich, S. Garcia, and C. Rodriguez. Effectiveness of highly active antiretroviral therapy in reducing heterosexual transmission of hiv. In *Anales Sis San Navarra*, volume 28. SciELO Espana, 2005.

[35] H.E. Cejtin, A. Kalinowski, P. Bacchetti, R.N. Taylor, D.H. Watts, S. Kim, L.S. Massad, S. Preston-Martin, K. Anastos, M. Moxley, et al. Effects of human immunodeficiency virus on protracted amenorrhea and ovarian dysfunction. *Obstetrics & Gynecology*, 108(6):1423, 2006.

[36] S.H. Chen, Y.C. Chen, and J.T. Liu. The effects of sibling sex composition on child mortality and education. 2010.

[37] Victor G Cicirelli. Sibling relationships in cross-cultural perspective. In *Sibling Relationships Across the Life Span*, pages 69–85. Springer, 1995.

[38] D. Coetzee, K. Hildebrand, A. Boulle, G. Maartens, F. Louis, V. Labatala, H. Reuter, N. Ntwana, and E. Goemaere. Outcomes after two years of providing antiretroviral treatment in khayelitsha, south africa. *Aids*, 18(6):887, 2004.

[39] Myron S Cohen, Ying Q Chen, Marybeth McCauley, Theresa Gamble, Mina C Hosseinipour, Nagalingeswaran Kumarasamy, James G Hakim, Johnstone Kumwenda, Beatriz Grinsztejn, Jose HS Pilotto, et al. Prevention of hiv-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, 365(6):493–505, 2011.

[40] N. Crepaz, T.A. Hart, and G. Marks. Highly active antiretroviral therapy and sexual risk behavior: a meta-analytic review. *Jama*, 292(2):224, 2004.

[41] UNAIDS World AIdS day report. How to get to zero: Faster smarter better. Technical report, UNAIDS, 2011.

[42] D. De Walque. How does the impact of an hiv/aids information campaign vary with educational attainment? evidence from rural uganda. *Journal of Development Economics*, 84(2):686–714, 2007.

[43] D. de Walque, H. Kazianga, and M. Over. Antiretroviral therapy awareness and risky sexual behaviors: Evidence from mozambique-working paper 239. *Working Papers*, 2010.

[44] J. Decosas, F. Kane, J.K. Anarfi, KD Sodji, and HU Wagner. Migration and aids. *Lancet*, 346(8978):826, 1995.

[45] E. Duflo, P. Dupas, and M. Kremer. Education, hiv and early fertility: Experimental evidence from kenya. 2011.

[46] E. Duflo, P. Dupas, M. Kremer, and S. Sinei. Education and hiv/aids prevention: evidence from a randomized evaluation in western kenya. *World*, 2006.

[47] P. Dupas. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *NBER Working Paper*, 2009.

[48] P. Dupas. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya, 2009.

[49] P. Dupas. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics*, 3(1):1–34, 2011.

[50] P. Dupas and J. Robinson. The (hidden) costs of political instability: Evidence from kenyaÕs 2007 election crisis. *California Center for Population Research, UCLA, On-line Working Paper Series*, page 1, 2010.

[51] Tim Dyson. The preliminary demography of the 2001 census of india. *Population and Development Review*, 27(2):341–356, 2001.

[52] D.K. Evans and E. Miguel. Orphans and schooling in africa: A longitudinal analysis. *Demography*, 44(1):35–57, 2007.

[53] T. Falola. *Teen life in Africa*. Greenwood Pub Group, 2004.

[54] D. Filmer, J. Friedman, and N. Schady. Development, modernization, and childbearing: The role of family sex composition. *The World Bank Economic Review*, 23(3):371–398, 2009.

[55] M. Fortes, AR Radcliffe-Brown, and D. Forde. African systems of kinship and marriage. *African systems of kinship and marriage*, 1950.

[56] O. Galor and D.N. Weil. From malthusian stagnation to modern growth. *The American economic review*, 89(2):150–154, 1999.

[57] Ashish Garg and Jonathan Morduch. Sibling rivalry and the gender gap: Evidence from child health outcomes in ghana. *Journal of Population Economics*, 11(4):471–493, 1998.

[58] M. Gersovitz, H.G. Jacoby, F.S. Dedy, and A.G. Tape. The balance of self-reported heterosexual activity in kap surveys and the aids epidemic in africa. *Journal of the American Statistical Association*, pages 875–883, 1998.

[59] J.R. Glynn, M. Caraël, B. Auvert, M. Kahindo, J. Chege, R. Musonda, F. Kaona, A. Buvé, et al. Why do young women have a much higher prevalence of hiv than young men? a study in kisumu, kenya and ndola, zambia. *Aids*, 15:S51, 2001.

[60] S.J. Goldie, Y. Yazdanpanah, E. Losina, M.C. Weinstein, X. Anglaret, R.P. Walensky, H.E. Hsu, A. Kimmel, C. Holmes, J.E. Kaplan, et al. Cost-effectiveness of hiv treatment in resource-poor settings–the case of cote d'ivoire. *New England Journal of Medicine*, 355(11):1141, 2006.

[61] M. Goldstein, J. Graff Zivin, J. Habyarimana, C. Pop Eleches, and T. Harsha. Aids treatment programs and sexual behavior. *Forum for Health Economics & Policy*, 2007.

[62] R.H. Gray, G. Kigozi, D. Serwadda, F. Makumbi, S. Watya, F. Nalugoda, N. Kiwanuka, L.H. Moulton, M.A. Chaudhary, M.Z. Chen, et al. Male circumcision for hiv prevention in men in rakai, uganda: a randomised trial. *Lancet*, 369:657–66, 2007.

[63] E.C. Green, D.T. Halperin, V. Nantulya, and J.A. Hogle. Uganda's hiv prevention success: the role of sexual behavior change and the national response. *AIDS and Behavior*, 10(4):335–346, 2006.

[64] J. Guyer. Dynamic approaches to domestic budgeting: cases and methods from africa. *A Home Divided: women and income in the Third World*, pages 155–172, 1988.

[65] J. Homsy, R. Bunnell, D. Moore, R. King, S. Malamba, R. Nakityo, D. Glidden, J. Tappero, and J. Mermin. Reproductive intentions and outcomes among women on antiretroviral therapy in rural uganda: a prospective cohort study. *PLoS One*, 4(1):4149, 2009.

[66] S. Jayachandran and A. Lleras-Muney. Life expectancy and human capital investments: Evidence from maternal mortality declines. *The Quarterly Journal of Economics*, 124(1):349, 2009.

[67] C. Juhn, S. Kalemli-Ozcan, and B. Turan. Hiv and fertility in africa: First evidence from population based surveys, 2008.

[68] Robert Kaestner. Are brothers really better? sibling sex composition and educational achievement revisited. *Journal of Human Resources*, 32(2):250–284, 1997.

[69] S. Kalemli-Ozcan. Does the mortality decline promote economic growth? *Journal of Economic Growth*, 7(4):411–439, 2002.

[70] S. Kalemli-Ozcan. A stochastic model of mortality, fertility, and human capital investment. *Journal of Development Economics*, 70(1):103–118, 2003.

[71] S. Kalemli-Ozcan. Aids," reversal" of the demographic transition and economic development: Evidence from africa. *NBER Working paper*, 2006.

[72] M.M. Kitahata, S.J. Gange, A.G. Abraham, B. Merriman, M.S. Saag, A.C. Justice, R.S. Hogg, S.G. Deeks, J.J. Eron, J.T. Brooks, et al. Effect of early versus deferred antiretroviral therapy for hiv on survival. *New England Journal of Medicine*, 360(18):1815–1826, 2009.

[73] J. Kluve, H. Lehmann, and C.M. Schmidt. Active labor market policies in poland: human capital enhancement, stigmatization, or benefit churning? *Journal of Comparative Economics*, 27:61–89, 1999.

[74] S.P. Koenig, F. Leandre, and P.E. Farmer. Scaling-up hiv treatment programmes in resource-limited settings: the rural haiti experience. *Aids*, 18:S21, 2004.

[75] J.L. Kornreich, K.D. Hearn, G. Rodriguez, and L.F. O'Sullivan. Sibling influence, gender roles, and the sexual socialization of urban early adolescent girls. *Journal of Sex Research*, 40(1):101–110, 2003.

[76] D. Lakdawalla, N. Sood, and D. Goldman. Hiv breakthroughs and risky sexual behavior*. *The Quarterly Journal of Economics*, 121(3):1063–1102, 2006.

[77] C. Laurent, N. Diakhaté, N. Fatou Ngom Gueye, M. Awa Touré, P. Salif Sow, M. Awa Faye, M. Gueye, I. Lanièce, C. Touré Kane, F. Liégeois, et al. The senegalese government's highly active antiretroviral therapy initiative: an 18-month follow-up study. *Aids*, 16(10):1363, 2002.

[78] A. Leibowitz, W.K. Chow, and M. Eisen. An economic model of teenage pregnancy decision-making. *Demography*, 23(1):67–77, 1986.

[79] A. Leigh. Trust, inequality and ethnic heterogeneity. *Economic Record*, 82(258):268–280, 2006.

[80] F.R. Lichtenberg. The impact of increased utilization of hiv drugs on longevity and medical expenditure: an assessment based on aggregate us time-series data. *Expert Review of Pharmacoeconomics and Outcomes Research*, 6(4):425–436, 2006.

[81] E. Lieberman. Ethnic politics, risk and policy making: A cross-national statistical analysis of government responses to hiv/aids, forthcoming in. *Comp. Polit. Stud*, 2007.

[82] B. Linkow. Fraying of the ties that bind: Hiv/aids and informal contract enforcement in kwazulu natal, south africa. 2009.

[83] C.B. Lloyd. *Growing up global: The changing transitions to adulthood in developing countries*. Natl Academy Pr, 2005.

[84] C.B. Lloyd and B.S. Mensch. Marriage and childbirth as factors in dropping out from school: an analysis of dhs data from sub-saharan africa. *Population Studies*, 62(1):1–13, 2008.

[85] C.B. Lloyd and UN Millennium Project. *Schooling and adolescent reproductive behavior in developing countries*. Millennium Project, 2006.

[86] J.R.P. Marins, L.F. Jamal, S.Y. Chen, M.B. Barros, E.S. Hudes, A.A. Barbosa, P. Chequer, P.R. Teixeira, and N. Hearst. Dramatic improvement in survival among adult brazilian aids patients. *Aids*, 17(11):1675, 2003.

[87] M. Marston, K. Harriss, and E. Slaymaker. Non-response bias in estimates of hiv prevalence due to the mobility of absentees in national population-based surveys: a study of nine national surveys. *Sexually transmitted infections*, 84(Suppl 1):i71, 2008.

[88] L.S. Massad, G. Springer, L. Jacobson, H. Watts, K. Anastos, A. Korn, H. Cejtin, A. Stek, M. Young, J. Schmidt, et al. Pregnancy rates and predictors of conception, miscarriage and abortion in us women with hiv. *Aids*, 18(2):281, 2004.

[89] S. Mechoulan. Risky sexual behavior, testing, and hiv treatments. *Economics*, 10(2):5, 2007.

[90] E. Miguel and M.K. Gugerty. Ethnic diversity, social sanctions, and public goods in kenya. *Journal of Public Economics*, 89(11-12):2325–2368, 2005.

[91] F.H. Milne and D.S. Judge. Brothers delay menarche and the onset of sexual activity in their sisters. *Proceedings of the Royal Society B: Biological Sciences*, 278(1704):417–423, 2011.

[92] V. Mishra, A. Medley, R. Hong, Y. Gu, and B. Robey. Levels and spread of hiv seroprevalence and associated factors: evidence from national household surveys. *Levels and spread of HIV seroprevalence and associated factors: evidence from national household surveys*, 2009.

[93] N. Monteiro. Using propensity matching estimators to evaluate the impact of privatisation on wages. 2004.

[94] J. Morduch. Sibling rivalry in africa. *The American Economic Review*, 90(2):405–409, 2000.

[95] D. Morgan, C. Mahe, B. Mayanja, J.M. Okongo, R. Lubega, and J.A.G. Whitworth. Hiv-1 infection in rural africa: is there a difference in median time to aids and survival compared with that in industrialized countries? *Aids*, 16(4):597, 2002.

[96] L. Myer, R.J. Carter, M. Katyal, P. Toro, W.M. El-Sadr, and E.J. Abrams. Impact of antiretroviral therapy on incidence of pregnancy among hiv-infected women in sub-saharan africa: A cohort study. *PLoS Medicine*, 7(2):4–26, 2010.

[97] Mary B Oliver and Janet S Hyde. Gender differences in sexuality: a meta-analysis. *Psychological bulletin*, 114(1):29, 1993.

[98] E. Oster. Sexually transmitted infections, sexual behavior, and the hiv/aids epidemic*. *Quarterly Journal of Economics*, 120(2):467–515, 2005.

[99] E. Oster. Routes of infection: Exports and hiv incidence in sub-saharan africa. 2007.

[100] E. Oster. Hiv and sexual behavior change: Why not africa? *Journal of Health Economics*, 2012.

[101] L. Ouyang. Sibling effects on teen risky behaviors. 2004.

[102] R. Pongou. Anonymity and infidelity: Ethnic identity, strategic cross-ethnic sexual network formation, and hiv/aids in africa. *Unpublished paper, Department of Economics, Brown University*, 2009.

[103] T.C. Porco, J.N. Martin, K.A. Page-Shafer, A. Cheng, E. Charlebois, R.M. Grant, and D.H. Osmond. Decline in hiv infectivity following the introduction of highly active antiretroviral therapy. *AIDS (London, England)*, 18(1):81, 2004.

[104] D.N. Posner. Measuring ethnic fractionalization in africa. *American Journal of Political Science*, 48(4):849–863, 2004.

[105] L.H. Pritchett. Desired fertility and the impact of population policies. *Population and Development Review*, pages 1–55, 1994.

[106] G. Reniers and J. Eaton. Refusal bias in hiv prevalence estimates from nationally representative seroprevalence surveys. *AIDS (London, England)*, 23(5):621, 2009.

[107] C.A. Sabin and A.N. Phillips. Should hiv therapy be started at a cd4 cell count above 350 cells/[mu] l in asymptomatic hiv-1-infected patients? *Current opinion in infectious diseases*, 22(2):191, 2009.

[108] J. Santelli, R. Rochat, K. Hatfield-Timajchy, B.C. Gilbert, K. Curtis, R. Cabral, J.S. Hirsch, and L. Schieve. The measurement and meaning of unintended pregnancy. *Perspectives on sexual and reproductive health*, 35(2):94–101, 2003.

[109] G.P. Schmid, A. Buvé, P. Mugyenyi, G.P. Garnett, R. J Hayes, B.G. Williams, J.G. Calleja, K.M. De Cock, J.A. Whitworth, S.H. Kapiga, et al. Transmission of hiv-1 infection in sub-saharan africa and effect of elimination of unsafe injections. *The Lancet*, 2004.

[110] Roberto Serrano and Roland Pongou. A dynamic theory of fidelity networks with an application to the spread of hiv/aids. 2009.

[111] D.J. Smith and B.C. Mbakwem. Life projects and therapeutic itineraries: marriage, fertility, and antiretroviral therapy in nigeria. *Aids*, 21:S37, 2007.

[112] AL Stangl, N Wamai, J Mermin, AC Awor, and RE Bunnell. Trends and predictors of quality of life among hiv-infected adults taking highly active antiretroviral therapy in rural uganda. *AIDS care*, 19(5):626–636, 2007.

[113] R.L. Stoneburner and D. Low-Beer. Population-level hiv declines and behavioral risk avoidance in uganda. *Science*, 304(5671):714, 2004.

[114] R.L. Thornton. The demand for, and impact of, learning hiv status. *The American economic review*, 98(5):1829, 2008.

[115] R.L. Trivers and D.E. Willard. Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179(4068):90–92, 1973.

[116] WHO UNAIDS. 2009 aids epidemic update. *Aids*, 10(5), 2009.

[117] J.L. Vigdor. Community composition and collective action: Analyzing initial mail response to the 2000 census. *Review of Economics and Statistics*, 86(1):303–312, 2004.

[118] Ingrid Waldron. Sex differences in human mortality: the role of genetic factors. *Social science & medicine*, 17(6):321–333, 1983.

[119] H.A. Weiss, M.A. Quigley, and R.J. Hayes. Male circumcision and risk of hiv infection in sub-saharan africa: a systematic review and meta-analysis. *Aids*, 14(15):2361, 2000.

[120] A. Whiteside. Poverty and hiv/aids in africa. *Third World Quarterly*, 23(2):313–332, 2002.

[121] B.G. Williams, D. Taljaard, C.M. Campbell, E. Gouws, L. Ndhlovu, J. Van Dam, M. Carael, and B. Auvert. Changing patterns of knowledge, reported behaviour and sexually transmitted infections in a south african gold mining community. *Aids*, 17(14):2099, 2003.

[122] R.J. Willis. A theory of out-of-wedlock childbearing. *Journal of Political Economy*, 107(S6):33–64, 1999.

[123] K. Wools-Kaloustian, S. Kimaiyo, L. Diero, A. Siika, J. Sidle, C.T. Yiannoutsos, B. Musick, R. Einterz, K.H. Fife, and W.M. Tierney. Viability and effectiveness of large-scale hiv treatment initiatives in sub-saharan africa: experience from western kenya. *Aids*, 20(1):41, 2006.

[124] A. Young. The gift of the dying: The tragedy of aids and the welfare of future african generations*. *Quarterly Journal of Economics*, 120(2):423–466, 2005.

[125] A. Young. In sorrow to bring forth children: fertility amidst the plague of hiv. *Journal of economic growth*, 12(4):283–327, 2007.